

Strategy needs disciplines

The UK government will soon reveal its thinking about universities. There are critical challenges in science to be tackled by researchers, funders and advocates acting more coherently within disciplinary contexts.

The forthcoming White Paper from the British government on the future of higher education will prove inadequate if it bases its recommendations on university research only on the advice of the Dearing committee. That committee, established to stimulate thinking on higher education, skated over some of the critical problems affecting the research community. Worryingly, there is little sign that the Labour government's Department for Education and Employment (DFEE) will be thinking any more deeply about research.

The financial problems are severe: the evidence of inadequate equipment is widely accepted. The Dearing committee's partial panacea — a substantial "revolving loan fund" — has already proved to be a mirage. The government must either find hundreds of millions of pounds or send a clear message that it is unable to do so for the foreseeable future. And, sooner rather than later, the DFEE (which funds university staff and infrastructure) and the Office of Science and Technology (funding the research councils) must together clarify criteria for funding that in effect signal the intended scale and diversity of the research community. Only then will the research and higher education community be able to plan effectively, and only then can tough decisions reshaping the research base be taken sensibly.

As was clear from a meeting about these issues last week, organized by *Nature*, many of those working with the university system are frustrated by the absence of a clear national strategy and also by the feeling that advocacy on their behalf is ineffective. Progress will best be made from the principle that strategy and advocacy are most powerful when directed at or from scientific disciplines. An individual's scientific identity and motivation are forged within a discipline, and teaching is sought and given on a disciplinary basis — hence the continuing importance of disciplinary departments. Research strategies of nations and of employers requires analysis by disciplines, whereas international networks develop within disciplinary frameworks. Although multidisciplinary approaches are required more and more, those tend to address specific problems on which strong

disciplinary skills are brought to bear.

Future planning needs to work with that grain, bestowing more effectively the rewards of peer-reviewed excellence given by funding councils to universities and by research councils, industry and charities to those who deliver the goods: researchers. The underlying goal for the government's funding bodies must be to maximize research productivity and the ability of departments and research groups to act as intellectual and educational entrepreneurs.

That necessitates streamlining and cross-referencing peer assessment across government agencies to minimize its burden and also more directly rewarding departments for their progress — even at the expense of some of the discretion of vice-chancellors over money from funding councils. University departments have been through three research assessment exercises within the past decade and have responded to research councils and industry as a result of national foresight exercises. Their research staff are now of unprecedentedly high calibre. Time, then, to set about diminishing the practical burdens of accountability, and extending the periods over which excellent scientists are guaranteed funding. And good working relationships and transparency between researchers and department heads will be ever more essential if departments are to act effectively.

What of advocacy to government? Universities have diverse agendas internally and externally, and will find it hard to improve their collective representation. Messages from Save British Science and the much more powerful Royal Society will inevitably be blunted by their need to represent all disciplines. Biologists have their case ready-made: molecular biology is of immediate relevance to national wealth and health, while the powerful UK pharmaceutical companies are acting coherently as a lobbying force for their discipline. Given the competition for funds, the need for the lobbying skills of the Institute of Physics and the Royal Society of Chemistry and associated industries has never been greater. The pressure of advocacy will increasingly fall on them, while departments adapt or die. □

Make radio polluters pay

Governments need to do more to curb pollution of astronomically essential radio bands.

In the film *Contact*, extraterrestrial life announced its presence to Jodie Foster using radio signals and the universal language of mathematics. It would be nice to think that the public, who flocked to see the film, is now sensitized to the cultural importance of radioastronomy, and might be dismayed to learn of its threatened obliteration by mobile telephones.

Pollution of radiofrequencies is perpetrated by commercial satellite companies and permitted by national governments. Their meeting point is the International Telecommunication Union (ITU), a United Nations body which allocates the radiospectrum and controls radio regulations. The ITU accommodates commercial needs by ignoring one of its own rules: "harmful interference should not be caused to radioastronomy". An unwritten qualification — "provided that it does not cost too much" — dominates decisions. Thus the ITU has agreed toothless and non-mandatory guidelines for regulating

the quality of satellite transmissions (see page 103). These are insufficient to stop overspill emissions.

In the same way that they enforce regulations to protect rare habitats and designated areas of scientific interest, national governments should respect their obligations to radioastronomy. They should instruct their ITU delegates — and the ITU is only as good as its delegates — to make no concessions to industrial lobbies that risk pollution of radio bandwidths of scientific importance. Yet they do not.

Individual governments also have power outside the ITU because they are responsible for issuing national licences. They should reject pending licence applications from Iridium, which wants to operate a multisatellite mobile telephone system with no guaranteed protection for radioastronomy frequencies. The customers of Iridium, and similar companies, must be forced in this way to pay for non-polluting systems. □



Astronomers fight incursion of phones into radio frequencies

[GENEVA] Radioastronomers fighting mobile telephone companies whose transmissions threaten to interfere catastrophically with reception of radio signals at low frequencies are facing a new battle.

The threat comes from several companies that are investigating the use of broadband satellite transmissions at higher frequencies. This would provide higher capacity mobile telephones or Internet services.

The four-week World Radiocommunication Conference-97 (WRC-97), currently taking place in Geneva, seems unlikely to deliver what radioastronomers want — guaranteed transmission standards that would assure no harmful overspill emission into parts of the spectrum allocated to another user. Instead, radioastronomers are likely to have to continue negotiating protection of their bandwidths on a case-by-case basis, a procedure in which they have so far mostly been losers.

But the WRC-97 conference, run by the International Telecommunication Union (ITU) — a United Nations body that rules on issues relating to sharing of the spectrum, radio regulations and operational standards — has at least made one decision in radioastronomers' favour. It has agreed to consider at its 1999 meeting the sharing of the spectrum above 70 GHz, the highest part of the radio spectrum, which has so far not been targeted commercially because of the expensive technology that would be involved.

Until now radioastronomers have enjoyed free use of this part of the spectrum, which contains a very high density of spectral lines corresponding to important molecules. Several existing facilities are active in this range, as will be the planned next-generation Square Kilometre Telescope.

But radioastronomers have learnt the wisdom of fighting for guaranteed protection before commercial interests appear. The new pressure on the radio spectrum means that companies are gearing up to reserve sections of the millimetre-range spectrum, even though there is as yet no technology to exploit it cost-effectively. This time around, radioastronomers want watertight rules on overspill transmission before companies make specific technical plans.

The radio spectrum was last carved up by the radiocommunication section of the ITU in 1979, and radioastronomy was allocated around two per cent of the whole spectrum. Lack of commercial competition meant that it was allocated 24 per cent of the spectrum above 70 GHz.

But commercial interests, particularly

mobile telephone companies, have mushroomed since then. Radioastronomers, accustomed to regular local battles to defend their frequencies against interference from terrestrial broadcasts, found themselves threatened with complete disablement by interference from satellite communications (see below).

The increasing sophistication of mobile phone satellite systems operating at frequencies below 30 GHz threatens to destroy the ability of radiotelescopes to receive important, but weak, radio signals that scientists need to study early events of the Universe.

Such weak signals could be drowned out completely. One standard mobile telephone

placed on the Moon would be the third strongest radiosource observable in the whole Universe, says Willem Baan, chairman of the International Council of Scientific Unions' Interunion Commission on the Allocation of Frequencies for Radioastronomy and Space Science.

Companies such as Telesed, owned by Microsoft, are keen to use broadband satellite systems at frequencies up to 60 GHz. But this would interfere with other important frequencies allocated to radioastronomy.

The danger comes from overspill from frequencies in which the companies operate into important radioastronomy frequencies, which are often adjacent. Such 'spurious

Interference threatens Indian radiotelescope



[NEW DELHI] India's radioastronomers are upset about the likely impact of emissions from a new mobile telephone satellite network on the US\$17 million Giant Metre Wave Radio Telescope (GMRT) near Pune (above). The telescope, operated by the Tata Institute of Fundamental Research, is one of the most sensitive in the world.

The astronomers are calling on the government to protect their interests against Iridium, the Motorola subsidiary, which owns the largest share of the global satellite communication system. The company is also building a 'gateway' at a site only 80 km from GMRT, to service mobile telephone users in most of south Asia.

Astronomers at the Tata Institute's National Centre for Radio Astrophysics in Pune are worried that signals from the satellites to mobile

telephones will interfere with observations of the radio emission from hydroxyl (OH) molecules — an important band for exploration of star-forming regions in the galaxies. "We cannot allow this to happen," says the centre's director V. K. Kapahi.

The telescope operates in the 38–1,427 MHz frequency range, and there are plans to extend this to 1,670 MHz. Once this is done, GMRT will provide a larger collecting area than that of the Parkes Radio Telescope in Australia for survey of the southern sky for OH emitting clouds, says Govind Swarup, creator of the project. "It would be a pity if we cannot use our best telescope to scan for OH masers — or look for signals from extraterrestrial intelligence."

The government has taken care not to erect any radio transmitters within 500 km of the telescope. Several

frequency bands of interest to GMRT have also been given protection by the national committee that allocates frequencies for various services. But all this will change when Iridium satellites go commercial next September, says Kapahi.

The satellites will beam signals to phones in the 1,621.35–1,626.5 MHz frequency range which is close to the radioastronomy band of 1,610.6–1,613.8 MHz. The International Telecommunication Union allows satellite companies to downlink at these frequencies provided no "harmful interference" is caused to radioastronomy, but they still require a licence from the countries served.

Astronomers want the government to put Iridium's licence on hold until they see the full results of tests of the Iridium downlink emissions. **K. S. Jayaraman**

emissions' could be eliminated or masked by technical fixes. But these would be costly.

Radiotelescope operators have refused to accept offers from companies such as the mobile-phone operation Iridium to solve the problem by time-sharing frequencies or by building special transmitting (and therefore potentially interfering) beacons at their sites which could instruct handsets in the vicinity to change frequencies (see *Nature* 380, 569; 1996).

Companies have rejected suggestions from radioastronomers to change their frequencies, particularly those of downlinks which are the biggest source of potential interference, develop filters for their satellites to block spurious emissions, or cap the number of subscribers making calls at any one time to ensure minimum overspill.

The deadlock continues at the WRC-97 meeting, where radioastronomers' pleas have cut little ice. WRC-97 is finalizing allocations in the 30–60 GHz range for broadband services such as the planned multisatellite system of Telesed and Alcatel, called Sky

Bridge, as well as Sky station. This is a proposed system of stratospheric balloons which would hang over cities beaming the Internet to their inhabitants, and providing larger capacity mobile telephone links.

Radioastronomers have been using the conference to seek greater protection for their bandwidths at these higher frequencies. "We are worried that these high density bands will be unable to clean up the signals they transmit to avoid radioastronomy frequencies," says Baan.

Radioastronomers are divided about how best to fight the powerful commercial interests against which they are pitted. Some feel that the market atmosphere of the ITU and its World Radiocommunication Conferences is not the right battlefield. Lobbying in this arena is important but has had little effect, says Roy Booth, director of the Swedish Facility for Radioastronomy in Onsala.

In contrast, Booth recently organized a joint declaration from directors of radioastronomy institutes committing them to a global coordinated effort to control the

threat from satellite transmissions.

Harvey Butcher, director of the Netherlands Foundation for Research in Astronomy, believes action by governments is the only way to guarantee protection of radioastronomy. "Users with no economic interests are not getting a look-in [at ITU]," he says.

He is taking part in a Megascience Forum on Radioastronomy set up by the Organization for Economic Cooperation and Development last year. He believes that the contacts this brings with government science policy-makers could help to encourage them to provide money for research into technical fixes, such as development of filters, or to establish radio-quiet zones around observatories which satellite companies would be required to observe.

Baan says the ITU relies on financial contributions from industry and has conceded too much to industry in its decision making process. But he says that abandoning the battle within the ITU would be disastrous. "Things for radioastronomers will get worse before they get better."

Alison Abbott

Congress moves swiftly to protect academy's independence

[WASHINGTON] The US Congress was poised at the beginning of this week to exempt the National Academy of Sciences (NAS) from the Federal Advisory Committee Act (FACA), ending a year of uncertainty about how the academy prepares scientific and technical advice for the US government.

In the early hours of Monday, the House of Representatives passed a bill allowing the exemption, while requiring the academy to take several steps to open up its processes to the public. The Senate was expected quickly to pass a similar bill, which President Bill Clinton is ready to sign into law.

Congress has responded with unusual haste after a Supreme Court decision last week left the academy complex facing restrictions which Bruce Alberts, the president of the academy, said would prevent it from functioning independently of the government.

Committees organized by the academy would be more open to public scrutiny under the planned legislation. But the National Research Council (NRC), the executive arm of the academy, would still appoint and manage its own study committees, which academy officials have always said is their top priority.

As expected, the Supreme Court had refused to consider an appeal by the NAS against the decision of a lower court forcing NRC committees to operate under the rules of FACA (see *Nature* 386, 309; 1997). Those rules include allowing public attendance at most meetings, and giving federal agency employees responsibility for the setting up and management of advisory committees.



Horn: 'broad agreement' on need for exemption.

Academy officials argued that placing NRC committees under the direct control of federal officials whose programmes are under scrutiny would compromise their independence.

Having lost in the courts, the academy has taken its case to

Congress. Stephen Horn (Republican, California), who chairs a House of Representatives subcommittee in charge of government management issues, was largely responsible for drafting legislation exempting the NAS from the act.

At a hearing last week, Horn said there was "broad agreement" in both houses of Congress and in the Clinton administration to grant such an exemption.

But the academy has promised to take some steps to open its operations to the public. The names, biographies and brief conflict-of-interest disclosure statements of each member of an academy study panel would be posted on the Internet.

The public would then have a period, perhaps 20 days, in which to comment. It would be "totally up to the academy's discretion" what to do with those comments, says William Colglazier, the NRC's executive director.

The names and biographies of external reviewers who evaluate completed NRC reports before they are released would also

be posted. At present, the reviewers are anonymous, and Colglazier admits that releasing their names might make it more difficult to recruit volunteer reviewers. The academy would encourage public attendance at committee fact-finding meetings (which are already open) by posting notices on the Internet. Meetings involving internal deliberations of a committee or report writing would remain closed.

Alberts said at last week's hearing that these measures would "let as much sunshine in as possible without changing the independent nature of the advice that we give".

Christopher Paine of the Natural Resources Defense Council (NRDC) calls the academy's concessions to openness "a step in the right direction". The NRDC successfully sued to prevent the Department of Energy from using an academy study on its planned National Ignition Facility earlier this year on the grounds that it had not operated under the act's rules (see *Nature* 385, 755; 1997).

But Paine would like to see all NRC meetings open to outside observers and to have greater public accountability in the composition of the council's committees.

Paine's group will release a report this week detailing what he calls an "egregious failure" in the academy's system of internal controls in connection with the report on the ignition facility.

According to the NRDC, the energy department disbanded its own advisory panel and solicited the academy report instead because it thought it might be more favourable to the facility.

Tony Reichhardt

Japan ties the industry/university knot

[TOKYO] Japan's Ministry of International Trade and Industry (MITI) and the Ministry for Education, Science, Sports and Culture (Monbusho) are considering setting up 'technology liaison offices' to help improve collaboration between universities and industry.

The plan is one of several recommendations in a proposed new bill for promoting industry/university collaboration, which the two ministries announced last week. The liaison offices would either be set up individually at universities, or exist as regional offices that could be accessed by several universities in the area.

In contrast to the United States, there is comparatively little interaction in research and development between universities and industry in Japan. This is particularly true at national universities, where civil service law prevents scientists from being involved in research in industrial laboratories or profit-making activities. Thus, whereas US universities obtained 1,900 patents in 1994, for example, Japanese universities gained only 130 in the same year, according to MITI.

The proposed bill will allow private companies to fund universities setting up venture businesses, in contrast to the current system, under which funding from companies to private and national universities is restricted to grants, scholarships and donations. Monbusho is also considering modifying the civil service law to enable scientists at national universities to help set up venture companies and to get involved in research in industry.

According to MITI, the liaison offices will assist universities with drawing up research contracts for joint projects with industrial laboratories, and will help companies to set up their own research centres at universities.

MITI's Masahiro Hashimoto is in charge of promoting interaction between industry and universities. He says the new bill, which will be submitted to the next session of the Diet (Japan's parliament) in December, will be a "very positive move".

Some national universities, such as Tokyo University, Tokyo Institute of Technology and Tsukuba University, have already set up liaison offices. But they have sometimes had to appoint postgraduate students as directors of these offices because of the law prohibiting university staff from involvement in profit-making activities.



"The technology liaison office can advise scientists on the commercial exploitation of inventions made from their research," says Hashimoto. "The new bill will allow national universities to profit from their inventions whereas any income gained from patents and industrial applications under the current system is collected by central government."

For industry and universities to collaborate successfully, the universities will have to change their attitude, says Susumu Nishimura, director of the research centre at Banyu Pharmaceuticals in Tsukuba science city. "Researchers at universities must change their perception that industry is just a source of funding. If there is to be 'real' collaboration between universities and industry, they (universities) will need to start implementing 'commercially applicable' research from which industry can benefit."

Nishimura adds that the present environment in Japanese universities will make it difficult for venture businesses to thrive. "There is still a strong feeling that good researchers do not normally get involved in business activities."

"Japanese universities and national research centres are not equipped with adequate measures to protect intellectual property and to advise on appropriate procedures for commercial development of inventions," says Nishimura. "It is quite often the case that researchers miss their chance to obtain patents from their inventions simply due to lack of support in that area."

Asako Saegusa

British study will assess risks of vCJD blood transmission

[PARIS] The British government last week launched a study into the risk that the agent that causes the new variant of Creutzfeldt-Jakob disease (vCJD) might be transmitted by blood, and instructed the National Blood Authority to look at ways of reducing the potential risk.

The moves were announced by Frank Dobson, the health secretary, following advice from the government's Spongiform Encephalopathy Advisory Committee (SEAC). The committee says recent research suggests that the pathogenesis of vCJD differs from that of classical CJD in that the causative agent may occur in particular in lymphocytes, and that it would therefore be "logical" to consider removing such cells from blood products.

SEAC pointed out that while it is so far impossible to assess the risk of transmission of vCJD, risk assessments are urgently needed to obtain a clearer picture. Dobson said his department will carry out a full study immediately, and will consider

introducing the leukodepletion of blood, depending on the outcome of the study.

Steve Dealler, a bovine spongiform encephalopathy (BSE) researcher at Burnley General Hospital, who has been campaigning vigorously about the potential risks of contaminated blood (see *Transfusion Medicine*, 6, 217; 1996), welcomed the government's action, but says that it should have been taken immediately after the risk of BSE passing to humans was acknowledged in March last year.

Dealler and others are sceptical about whether leukodepletion of blood will be sufficient to eliminate all risk, as plasma is likely to be contaminated by leakage from broken cells. The solution would be to screen donors for the disease, he points out, arguing that this makes development of a diagnostic test for vCJD a priority.

Efforts to do so may be helped by the recent discovery of a monoclonal antibody specific for the abnormal version of the prion protein thought to cause the

disease (see *Nature* 390, 74; 1997).

The government "is acting responsibly" in taking a precautionary stance, says Tony Wilson, chief executive officer of the UK Haemophilia Society. But he remains concerned that economic factors may be given excessive weight in the cost-benefit analysis of the safety of blood products. He points out that most UK haemophiliacs are still treated with clotting factors from natural blood, which are inherently more risky than the recombinant alternatives widely used elsewhere, because the natural products are cheaper in Britain.

Meanwhile, the French national bioethics committee last week concluded that while the risk of transmission of vCJD by blood remained "hypothetical", a body should be set up to monitor scientific knowledge of the risk. Dealler adds that the renewed attention on the risk of blood products suggests it is only a matter of time before European countries raise the question of a ban on exports of British blood products.

Declan Butler

Greenhouse talks fail to smoke out US Kyoto strategy



[TOKYO & LONDON] Two days of separate talks between environment ministers of developed and developing countries ended in Tokyo on Sunday (9 November) having achieved little tangible progress towards legally binding

targets for greenhouse gas emissions.

Members of the European Union signalled that their proposal to reduce greenhouse gas emissions by 15 per cent by the year 2010 would be negotiable at the climate convention's conference next month in Kyoto. But, despite repeated efforts, the US delegation refused to be drawn on whether its own proposal to stabilize emissions at 1990 levels was similarly 'flexible'.

A parallel meeting of developing countries made virtually no headway. Only two ministers attended, and there were no delegations from China, India or Zimbabwe — some say because they were unhappy at being segregated from developed countries. "They would have preferred a meeting of all parties on the United Nations model," said one participant.

John Prescott, Britain's deputy prime minister, who chaired the developed country discussion, says he was pleased that countries appear to be moving away from rhetoric. "Every nation and everyone expressed that view," he says. "I think that is a single step forward, because up to now an awful lot of people have been predicting that, whatever the circumstances, they would not cooperate. When you have got a wide difference between zero and fifteen you have got to feel your way through to what is likely to be a realistic figure."

Delegates spent much of the time exploring whether the United States would be prepared to sign up to a higher target — for example a 10 per cent reduction in emissions from 1990 levels — achieved through 'joint implementation' and emissions trading.

Again, US delegates would not say whether this proposal was acceptable; according to one European Union participant, it was clear that the United States did not come to Tokyo prepared to talk in detailed terms.

Prescott is to hold separate bilateral meetings with India, Australia and New Zealand. The prime minister of Japan, which chaired the developing country meeting, held negotiations with King Fahd and Crown Prince Abdullah during a visit to Saudi Arabia. **Asako Saegusa & Ehsan Masood**

Russian parliament ratifies chemical weapons treaty

[MOSCOW] The Russian parliament last week ratified the United Nations convention on chemical weapons, originally signed in Paris in January 1993, despite opposition from critics who complained that Russia cannot afford the costs of destroying its stock of weapons. Soon after, the decree formalizing ratification was signed by President Boris Yeltsin.

Ratification had been actively supported in the Duma, the lower house, by the committee on international affairs and its chairman, Vladimir Lukin, who said that his position was shared by most members of the security and defence committees. The leaders of the factions and deputies' groups — including communists and the Liberal Democratic party — also supported the move.

Igor Ivanov, first deputy minister of foreign affairs, warned that refusal to ratify the convention would be used by Russia's enemies "whose number is still great" as one more reason for the North Atlantic Treaty Organization to move eastwards. Sanctions against Russian exports of chemical products would follow, as well as reduced foreign investment in its chemical industry.

Anatoly Kvashnin, the general staff commander, argued that the greatest threat from Russia's chemical weapons is now to the country itself, rather than its potential enemies, as the storage time of the weapons has long expired. "We need to destroy them to avoid an ecological catastrophe," he said.

But strong opposition to ratification came from the Duma's committee on industry, building, transport and power engineering. Stepan Sulakshin, its deputy chairman,

who was for many years a scientific worker at the Siberian branch of the Russian Academy of Sciences (RAS) in Tomsk, said that destroying all the chemical weapons in Russia would cost 34,400 billion rubles (US\$6 billion).

In 1994 and 1995, when Russia started to meet a presidential commitment to liquidate its chemical weapons, only 9.8 billion and 26.9 billion rubles respectively had been allocated to the task, and in 1996 there was no financing at all. In 1997 only 163.8 billion rubles was allocated.

The proposed budget for 1998 suggests that the equivalent of \$100 million, rather than the \$500 million needed, will be allocated. "Russia has no money to destroy its chemical weapons in 10 years, as demanded by the convention," Sulakshin told the Duma. "Foreign countries promised to consider supplying us with a little over \$100 million after the convention was ratified by Russia, but no agreement has been signed. Under these conditions, the ratification of the convention will ruin Russia."

He also said that Russia lacks reliable technology for destroying chemical weapons. Until now, the maximum weight of chemical ammunition destroyed at the laboratories was less than 50 grams, while more than 40 tonnes are to be liquidated. The destruction programme has also failed to pass environmental criteria, even though four Duma committees had insisted on this.

"The RAS suggestion to use the economic physical methods of plasma destruction was not adopted," Sulakshin said. But the Duma rejected his arguments and gave its support to the convention. **Carl Levitin**

US panel proposes joint 'pathogens initiative'

[WASHINGTON] The US Department of Defense (DOD) should back a collaborative research programme with Russia on dangerous pathogens, says a committee of the National Academy of Sciences (NAS) chaired by Nobel laureate Joshua Lederberg of Rockefeller University, New York.

The panel says that such a programme would enhance US national security by engaging in joint projects Russian scientists who formerly worked on biological weapons research.

Its findings would lead to extensive collaboration between biologists at Biopreparat, the body that ran the Russian biological weapons programme, and their counterparts at the US Centers for Disease Control, the US Army Medical Research Institute of Infectious Diseases and US universities.

The committee proposes that the DOD should spend \$38 million over five years on a programme called the Pathogens Initiative, with more than half of the money

going to support work in Russia.

The programme would also help the United States to ensure that Russia had stopped all work on biological weapons. Jo Husbands, director of NAS's Committee on International Security and Arms Control, says it is hard to ensure that research on infectious diseases cannot be applied militarily. "But the panel felt that the best insurance we can have is to be engaged with the [Russian] scientists," she says. **Colin Macilwain**

Malaysia backs 'gag' on haze scientists

[LONDON] The Malaysian government has justified a decision forbidding scientists at public institutions to talk to the press about the haze from Indonesian forest fires. It claims that a circular sent to vice-chancellors was designed to prevent scientists from publicizing the results of research before they had been peer-reviewed.

"That does not amount to a gagging order," says Zaini Ujang, a special adviser in the ministry of education. "Research on the haze is not comprehensive. There are many gaps, and we feel that it is not right for scientists to start talking to the press when many have only just begun to investigate this issue."

Parts of Malaysia remain covered by the haze, although pollution is slowly returning to 'normal levels'. Ujang's comments are widely seen as an effort to stem damage triggered by an interview by the Malaysian education minister, Najib Tun Razak, last week.

Scientists reacted with concern when the minister told reporters that the cabinet had barred scientists from making "negative statements" about sensitive issues such as the haze, on the grounds that such "speculative findings" were damaging the country's image abroad and harming tourism.

"Though some of them are experts in their respective fields, their statements on specific issues could give a worrying picture, and we want to avoid the alarmist attitude,"



Not a blanket ban: the guidance to scientists applies only to the haze, says the government.

the minister told the *Daily Star* newspaper. Scientists should defer to "higher authorities" before making public statements.

The next day, Razak told reporters the gagging order applied to the haze only, and was not a "blanket ban" on all issues. "We are trying to prevent statements which are not supported by concrete scientific evidence and could easily damage the image of the country," he said.

Three Malaysian government departments — science, education and environment — recently set aside 4 million ringgit (US\$1.2 million) for research into the health

effects of the haze. Ujang says that officials in the departments began to get annoyed when scientists on government research programmes made what he calls "conclusive statements" about the haze. One told a local newspaper that breathing the haze had the same effect as smoking 40 cigarettes a day.

"As funders of this research, they have a right to be annoyed," says Ujang. "When I was a research scientist with ICI in the UK, I had to brief the company before publishing in journals. Not all of my work was published. In fact, to this day some of my work has still not been allowed to be published."

As education minister, Razak lies third in the ruling party's pecking order, and is certain to challenge the deputy prime minister, Anwar Ibrahim, when the prime minister's job falls vacant. Some believe that Razak's comments will not only put back his promotion prospects but could also damage Malaysia's attempts to persuade scientists to communicate with the public.

"Scientists in Malaysia are notoriously wary of talking with the press or public," says Merryll Davies, editorial adviser to Malaysia's TV3 channel. "If I'd ring someone up for an interview they would say, 'have you spoken to my boss?' There's a strong culture of 'keep your head below the parapet and not be noticed'. The minister's comments will only encourage these habits."

Ehsan Masood

Largest-ever study contests radiation role in childhood cancers

[LONDON] The theory that men who have been exposed to radiation father children with a higher-than-average risk of contracting childhood leukaemia/non-Hodgkin's lymphoma (LNHL) received a new blow last week with the publication of the results of the largest study of its kind.

The survey used data from more than 100,000 workers occupationally exposed to radiation and confirmed a higher incidence of the disease among their children. But it did not establish a dose-response relationship. This suggests a factor other than radiation is the explanation.

Louise Parker, senior lecturer in epidemiology at the University of Newcastle upon Tyne, says the results, published in the *British Medical Journal* (7117, 1181-1188; 1997), "put another nail in the coffin" of the radiation theory, as did an earlier study published almost four years ago (see *Nature* 367, 678; 1994).

"I think the idea that preconception irradiation of the father is a cause of childhood LNHL is now a dead duck," says Leo Kinlen, director of the Cancer Epidemiology Research Group at the University of Oxford, and one of the paper's authors.

Controversy about whether paternal irradiation places children at a higher risk of LNHL has raged since the publication of a paper in 1990 by the late Martin Gardner of the University of Southampton. Gardner found children of workers exposed to radiation were about eight times more likely to succumb to LNHL than the general population (see *Nature* 343, 676; 1990).

The correlation of preconception irradiation with the increased risk of cancer in children subsequently became known as the 'Gardner hypothesis'. The finding surprised the scientific community, because similar conclusions could not be drawn from data from Hiroshima and Nagasaki.

Gerald Draper, principal author of last week's study, says that Gardner "did a good piece of epidemiology that generated a good hypothesis". But subsequent studies "have shown the hypothesis to be wrong".

Draper's study, explicitly set up to test the Gardner hypothesis, included all children up to 15 years of age born and diagnosed in Britain as having cancer between 1952 and 1986. The team matched the known childhood cases with individuals on the 120,000-strong British register of

workers who are regularly monitored for radiation exposure.

The research team found that, overall, workers on the register were indeed more likely than the general population to have children who developed cancer; the average risk by age 15 in Britain is 6.5 per 10,000 children, and the authors estimate that this increases by 5.4 per 10,000 among children of workers on the register.

But they also found that there was no dose-response relationship. In fact there was a greater risk of childhood cancer among children of fathers on the register who had received a zero dose of radiation.

While appearing to overturn the Gardner hypothesis, the study has nevertheless made a significant contribution to the study of childhood cancers. One explanation, championed by Kinlen, is that infection is passed between people coming into an area and the previously isolated population.

The study is one of two recommended by the Committee on the Medical Aspects of Radiation in the Environment in the wake of Gardner's publication. The second, the Nuclear Industry Family Study, is expected to be published next year.

Helen Gavaghan

'Mind body' claims put NIH on defensive

[WASHINGTON] The US National Institutes of Health (NIH) has been forced to defend its peer-review system in Congress against charges of bias against 'mind body medicine'. At the same time last week, it was hosting a conference that declared acupuncture a *bona fide* treatment for some kinds of nausea and pain, sparking criticism from scientific sceptics.

The charges about 'mind body medicine' — the theory that emotions are critical to human health and disease — and the questioning of the NIH's commitment to it, were made at a congressional briefing convened by Representative John Porter (Republican, Illinois), who holds key legislative responsibility for NIH's budget.

Porter argued at the briefing that traditional medicine "has seemed to have left out some very efficacious approaches". He added later that the NIH had to "give some real guidance to the American people" about what alternative treatments were safe and efficacious. He said: "I don't think [it is] doing that sufficiently."

The main target of criticism at the briefing was Steven Hyman, director of the National Institute of Mental Health (NIMH), who was assailed by supporters of 'mind body medicine' for what they described as an inveterate bias against their approach by NIH and its reviewers.

"The NIH has dug in its heels and hardened its position against mind body medicine," said Candace Pert of Georgetown University in Washington DC, a pharmacologist who discovered opiate receptors. But, she said, "people want it, it works, and it really does have scientific rationale".

James Gordon, director of the Center for Mind-Body Medicine at Georgetown, argued that there is a bias built into peer-review committees. He claimed that thousands of carefully controlled studies are published in peer-reviewed journals on many alternative therapies, but NIH peer reviewers are ignorant of them.

Hyman said review panels are groups of human beings and it is impossible to free them from all bias. But he could think of no better way of producing valid, objective results and therapies "than subjecting our proposals to do research to a community of peers in the marketplace of ideas". He told the briefing: "There have to be appropriate standards of evidence."

Although Porter, who chairs the House of Representatives spending subcommittee that funds NIH, does not have authority to write laws directing NIH to pay more attention to how the mind influences disease and health, he suggested that other committees with jurisdiction should take up the issue.

In addition to a "major" role in the area

for NIMH, Congress should question whether the Office of Alternative Medicine (OAM) and the Office of Behavioral and Social Sciences Research "are really fulfilling the expectations that Congress had" for them in this area, he said.

Porter said he was drawn to the issue in part because this summer he defied the recommendations of three surgeons that he have an operation for a slipped intervertebral disc. A month of rest and relaxation cured him, he says, and provoked his interest in the power of the mind to heal.

Meanwhile, an independent panel convened by the NIH in Bethesda, Maryland, declared at the end of a three-day 'consensus conference' that it had found "clear evidence" that acupuncture works for post-operative dental pain and for nausea and vomiting caused by chemotherapy, after surgery and probably during pregnancy.

The findings of the 12-member panel were immediately criticized by sceptics who claimed that the panel was "fixed" against

critics of acupuncture. "It was a conference of believers [who] recited their delusions as facts," said Victor Herbert, a professor of medicine at the Mount Sinai School of Medicine in New York, who dismisses acupuncture as "pseudo-religious cultism".

But the organizers of the conference, cosponsored by OAM and the Office of Medical Applications of Research, vigorously defended the panel and the process as fair.

Its conclusions were based largely on a three-month review of the available literature, which was judged by the "highest standards", said David Ramsay, the panel chair and president of the University of Maryland, Baltimore, and a former editor of the *American Journal of Physiology*. He said: "We did find positive evidence for the efficacy of acupuncture in certain conditions."

Alan Trachtenberg, a physician at NIH who chaired the conference planning committee, strongly denied any bias in the selection of 25 people to present data to the panel.

Meredith Wadman

India drafts law to protect bioresources

[NEW DELHI] The Indian government is planning tough legislation that will severely restrict access by foreign researchers and pharmaceutical companies to the country's biological resources.

Under the legislation, to be introduced in parliament next month, Indian scientists who transfer commercially valuable research results to foreign nationals without official approval will be punished.

The draft legislation has been drawn up by an expert committee headed by the eminent agricultural scientist M. S. Swaminathan, who says that India's main concern is to protect its biological resources from being exploited by foreigners without sharing the benefits with local people.

All political parties have welcomed the move to protect India's bioresources through the legislation, which has been worked on since India became a party to the 1993 Convention on Biological Diversity, which enables countries to safeguard their rights over their biological diversity.

Suman Sahai, a member of the expert committee and president of Gene Campaign, a nongovernmental organization, says the law will give India a legal basis on which to deal with the theft of genetic resources. "Right now, our biological resources are being plundered," she says.

The government is to set up a National Biodiversity Authority (NBA) to implement the law. It is to be headed by an eminent scientist, and will include representatives from the departments of environment,

agriculture, biotechnology and ocean development, as well as five people from outside the government. State governments will set up their own biodiversity boards.

Under the legislation, a non-Indian will be prohibited from "obtaining any biological resources for research or commercial utilization or collecting samples or undertaking any activity in the nature of bioprospecting without previous approval of the NBA". The law will make it illegal for an Indian citizen to transfer "the results of any research with respect to any biological resource for monetary consideration to any person who is not a citizen of India without NBA approval". Violators will face a five-year jail term and a fine of US\$30,000.

But the law will allow the transfer of biological resources to institutions abroad under government-sponsored collaborative research projects "subject to overall policy guidelines of central government". All agreements entered into before the act is passed "will be reviewed to bring them in conformity with the policy guidelines".

A key provision of the legislation is that no-one will be able to apply for a patent based on research or information gathered from any Indian biological resource without informing the NBA. Before giving permission for patenting, the authority will impose a benefit-sharing fee or royalty which will be credited to a biological diversity fund to be used to develop the communities which helped to conserve the biological resources.

K. S. Jayaraman

Australia urged to revise Antarctic efforts

[SYDNEY] The first comprehensive review of Australia's research in Antarctica in the 50 years since a permanent national programme was established has recommended sweeping changes. The aim is to divert funds to support more science if savings can be made in infrastructure costs.

The report, released last week by the Antarctic Science Advisory Committee after an exhaustive analysis of possible changes and uncertainties 30 years ahead, is "a trail-blazer" in the history of Antarctica, says the committee chair, Michael Stoddart, deputy vice-chancellor at the University of New England in Armidale, New South Wales.

Stoddart, a zoologist, points to a shift in government policy for Antarctica. Unlike all other areas of publicly supported research, he says, the direction of Australian science in Antarctica "is entirely in support of government policy positions". In its brief to the review, the government advanced its policy from "maintaining" the nation's interests in Antarctica to "enhancing" them.

In the complex territorial and, increasingly, commercial affairs of Antarctica, the review says that more science is central to "projecting an image of involvement with its Antarctic Territory" and supporting Australia's goal of exerting greater influence among the countries of the Antarctic Treaty System. Australia is bidding to have an office for the treaty established in Hobart, Tasmania, but agreement has yet to be reached on whether there should be such an office, let alone its location — Argentina also wants it.

Australia's work at four permanent bases and two part-time stations on the Antarctic continent and sub-Antarctic islands has become constrained by dependence on the transport of personnel, equipment and supplies solely by ice-breaking ship in summer months. Demands for more fishing and tourism, changes to sovereignty over land and

sea, and international collaboration have become major factors determining how research can be developed while government finance continues to be limited.

The review says savings could result from concentrating effort at Davis, one of Australia's three permanent bases on the continent, and converting Mawson and Casey bases to unstaffed, automated operation in winter months once technology makes this possible. Leasing or sharing one or two bases with other countries are also options.

As well as studying scientific, international and commercial interests and logistics, the committee commissioned a panel to study the intrinsic value of Antarctica. The main scientific goal identified by the committee is boosting the base of knowledge about Antarctic weather systems. Rex Moncur, director of Australia's Antarctic Division, agrees that information about the Antarctic region is "the weak link" in current models of global change.

Australia spends \$A60 million a year (US\$42 million) on the Antarctic Division, which employs about 120 scientists and 185 support staff full-time. The division supports a larger number of researchers from universities and research agencies for short periods of work in Antarctica.

Australia's year-round research and weather monitoring have to be serviced by the leased ice-breaker, *Aurora Australis*, which takes a month for a round trip. Eighty per cent of the division's budget is required for infrastructure, with about \$A20 million consumed by the ship charter and by helicopter operation close to the bases.

As the present supply ship can provide only 60 days of science at sea, the report advocates chartering a smaller research vessel for dedicated surveys of the Southern Ocean. The committee also highlights the need for greater flexibility. Australia is now



Time for retirement? Dependence on the *Aurora Australis* icebreaker is constraining research.

virtually alone among Antarctic countries in its total dependence on sea transport.

Building a full-sized airstrip at one base for intercontinental flights and small landing strips for light aircraft at other bases would allow scientists and equipment to be flown from Australia and around the territory for short periods of field work. Stoddart says: "We need to get researchers to where the science needs to be done, rather than around where the bases are placed."

The committee had no brief to cost options for ski-equipped aircraft able to use ice runways and conventional aircraft needing gravel runways. But Moncur says they range from \$A5 million to \$A15 million.

Building facilities for air transport would have environmental effects certain to stir activists campaigning to preserve Antarctica as wilderness. Moncur, whose division manages Australia's Antarctic territories and bases as well as the research, considers that there are incentives to share resources with countries keen to be involved in Antarctica, especially some of Australia's Asian neighbours.

Following the aim of researching global change, the committee gives priority to studies of the Southern Ocean and on the high plateau of the continent, where Australian astronomers are site-testing with Americans for a possible infrared telescope (see *Nature* 385, 194; 1997), environmental protection, meteorology, fisheries and tourism.

Any extension of tourism from the short ship-based visits of today to catering, for example, to "adventurers" would require at least a summer base to be available. The report concludes tourism would "best be undertaken away from research sites" and would need to stand alone and have significant capacity for search and rescue.

Ian Macdonald, parliamentary secretary responsible for Antarctic matters, welcomed the report. Rapid decisions are expected in time for implementation in the May 1998 budget, the last before Australia's election. The report is accessible on the Internet at www.antdiv.gov.au/foresight. **Peter Pockley**

'Disappointing lack of action' over illegal fishing

[SYDNEY] Sensitivities about exploiting Antarctica's resources surfaced at the meeting in Hobart, Tasmania, last week of the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR). This followed revelations of illegal fishing, mostly unreported, in the Southern Ocean. Rex Moncur, director of the Antarctic Division, estimates the poaching is worth A\$500 million (US\$350 million).

After Australian fishermen

and scientists expressed alarm at the rapid depletion of some species, notably the Patagonian toothfish, the Australian Navy apprehended foreign vessels alleged to be operating illegally in the zone around the Heard and Macdonald Islands.

The two ministers in the Australian government concerned with Antarctica and resources expressed disappointment that "more concrete action could not be agreed" by CCAMLR after

Australia pressed for other countries to follow its example of scientific observers on vessels to monitor catches and report vessels seen poaching.

But CCAMLR, which operates by consensus and has no direct powers to police its decisions, agreed only to prohibit fishing for toothfish in three areas for which conservation measures are not in place, and to "tougher" controls on port access. **P.P.**

Independent panel will oversee US research on Gulf War illnesses

[WASHINGTON] President Bill Clinton is appointing an independent, nongovernment panel to oversee research by the US Department of Defense (DOD) into the relationship between possible chemical and biological exposures and the ill-defined collection of maladies known as Gulf War illnesses. Clinton said in a statement that retired senator Warren Rudman would head the panel. Its other four or so members have yet to be named and its mandate will last at least a year.

The move is a response to a final report Clinton received on 31 October from his Presidential Advisory Committee on Gulf War Veterans' Illnesses (PACGWI), which calls for the DOD to be stripped of authority for overseeing research on chemical and biological exposures and their relationship to Gulf War illnesses (see *Nature* 390, 4; 1997).

The PACGWI report said that the Pentagon's investigation had been hampered by "an institutional culture and pervasive inclination" to dismiss evidence of possible chemical exposure of US troops. It also sharply criticized the Pentagon for funding studies that had not been competitively reviewed, and urged that it immediately stop.

Clinton will also ask the National Academy of Sciences to monitor continuing research on associations between Gulf War service and current illnesses for the purposes of future research and disability compensation.

First UN text on science and human rights

[PARIS] Almost 50 years after the adoption of the 1948 United Nations Universal Declaration of Human Rights, the 186 member states of Unesco were this week scheduled to adopt the first UN text governing science and human rights, the "universal declaration on the human genome and human rights".

The text consists of a set of broad principles affirming the rights and liberties of individuals against genetic discrimination, while upholding the principle of scientific liberty.

As *Nature* went to press, it appeared that the only prominent absences from the signatories would be Canada, which feels the text should include more specific references to the rights of the handicapped, ethnic minorities and women, and Israel, which opposes an article banning human cloning, arguing that it is inappropriate to include detailed provisions governing medical and research practices.

As a declaration, the text is not legally binding on signatories. But the possibility that it may be developed into a treaty cannot

be discounted, says Nöelle Lenoir, who chairs Unesco's International Bioethics Committee (IBC). The declaration confers on the IBC the role of overseeing the implementation of the text's provisions, and the IBC hopes to advise countries on setting up ethics committees.

Brain science centre opens in Japan

[TOKYO] Japan's Institute of Physical and Chemical Research (RIKEN) has opened a Brain Science Institute that will play a central role in an important Japanese initiative to develop research on the brain. The institute, headed by Masao Ito, one of Japan's leading neuroscientists, is part of a plan by the government to spend ¥2,000 billion (US\$16 billion) on a wide range of neuroscience over the next 20 years (see *Nature* 382, 105; 1996).

At the opening ceremony on 10 November, Akito Arima, president of RIKEN, expressed delight at the speed with which the new institute had been established, and said that this had been due to a successful combination of a "bottom-up approach" by scientists and a "top-down" approach by government and politicians who have recently become strong supporters of science.

BSE report sees progress on EU science advice

[PARIS] The European Parliament last week wrapped up its investigation of the bovine spongiform encephalopathy (BSE) crisis. The conclusion of the final report by the temporary committee on BSE is that the European Commission (EC) has made substantial progress in reforming its system of providing scientific advice. But the report "deplors" that although most of the EC officials responsible for handling BSE have been assigned to other duties, none has been otherwise disciplined.

The conclusion lifts the parliament's earlier threat to sack the EC's commissioners if its demands were not met. Some observers argue that the parliament has been highly successful in using the threat to obtain changes in the way the EC operates, in particular a radical reform of its scientific committees to distance them from economic interests and make their deliberations more open (see *Nature* 385, 664; 1997). Details of the committees and their proceedings are on <http://europa.eu.int/en/comm/spc/spc.html>.

CERN continues search for director-general

[GENEVA] The committee of council of CERN, the European Laboratory for Particle Physics, has once again failed to select a new director-general to succeed Christopher Llewellyn Smith, whose term of office expires at the end of next year. At a special

meeting last week, the committee managed to reduce the three-man shortlist to a single candidate, Albrecht Wagner, director-general of research at DESY, Germany's national research centre for particle physics.

But it also added a last-minute candidate, Luciano Maiani, head of INFN, Italy's National Institute for Nuclear Physics, who is currently also president of CERN council. A final decision must be made by CERN council at the end of December. Should Maiani be selected, the same meeting would have to select a new president of council.

Brookhaven reactor faces further delay

[WASHINGTON] The US Department of Energy says it will not be able to name a new contractor for the Brookhaven National Laboratory this month, as promised, because of a new law requiring it to give the Congress 60 days' notice before awarding such a contract. Officials say the delay will, in turn, push back by at least two months any decision to reopen the High Flux Beam Reactor.

A tritium leak from a storage tank at the reactor led the department to fire the existing contractor, Associated Universities Inc. (AUI), in May (see *Nature* 387, 114; 1997). The latest delay leaves the troubled laboratory in limbo with an acting director and AUI still in charge. The energy department has received two bids to run the laboratory, one from the State University of New York at Stony Brook in partnership with Battelle, and another from the Illinois Institute of Technology and Westinghouse.

Global change expert nominated for Scripps



[SAN FRANCISCO] Charles F. Kennel (left), an expert in global change and global observation systems, has been nominated to be director of the Scripps Institution of Oceanography in La

Jolla, California. It is part of the University of California, San Diego, and has a \$103-million budget, 300 research programmes and a four-ship research fleet.

Kennel has been executive vice chancellor since 1996 at the University of California, Los Angeles, where he has worked for 30 years. He served as associate administrator for the National Aeronautics and Space Administration's Mission to Planet Earth Program from 1994 to 1996, and won the 1997 James Clerk Maxwell Prize for contributions to basic plasma physics and plasma astrophysics. The UC regents are expected to consider Kennel's nomination at their January meeting.

Dangers of hyperbole in climate prediction

Sir — As one who believes that restrictions on carbon dioxide emissions will eventually be essential, I am alarmed by the dangers of hyperbole with which Tim O’Riordan lards his review of Ross Gelbspan’s book (*Nature* **389**, 685; 1997). He urges all to read the book to understand how Washington lobbyists have determined the US position at the Kyoto conference next month; his review will be taken by many as the proof for which they have been waiting of the calumny that the research community is both arrogant and politically naïve.

Lobbyists in the United States lobby for all possible causes, and have names such as the Federation of American Scientists, Friends of the Earth and Greenpeace. They abound because no US legislation is enacted without wide consultation by the Congress, invariably in public.

I share the opinion that this arrangement is preferable to the mostly private consultation that precedes legislation in European democracies other than in Scandinavia. Lobbying has not prevented the US Congress from enacting some of the world’s toughest legislation on air and water quality and prescription drugs — invariably in the teeth of opposition from lobbyists.

O’Riordan goes on to say that “scientists are trained not to transgress into the world of judgement and political bickering” — but then does just that himself with phrases such as “the United States seems politically and ideologically incapable of coming to

terms with the moral and inequitable aspects of global climate change”. He predicts that Kyoto will disappoint “the wishes of the scientific community”, but promises (threatens?) a “more aggressive and politicized science” afterwards.

This view of science does scant justice to admirable groups (lobbyists?) such as Pugwash, while the implication that Kyoto would be a done deal were it not for the obscurantism of the United States is an over-simplification, to put it mildly.

There are three reasons for regarding the Kyoto process with suspicion:

- (1) the predictions of the Intergovernmental Panel on Climate Change (IPCC) are persuasive as to the direction of temperature change but may well exaggerate the rate of change by a factor of two;
- (2) nothing in the IPCC volume on “impacts” suggests that there has yet been a serious study of the effect of the wished-for restrictions on the global economy or of how best holistically to manage a transition to a stable greenhouse; and
- (3) the problem of inequity (between rich and poor countries) is dealt with only crudely in the draft convention — at the very least, it should include agreed criteria for telling when developing countries become developed.

John Maddox

9 Pitt Street,
London W8 4NX, UK

Alphabetical listing

Sir — Discussion of the relationship between citation rates and the place of one’s name in the alphabet tends to use English-language publication. Different people use different alphabets. The initial of my surname is the third letter of the Russian alphabet, but is close to the end of the English one.

And there are different national traditions in compiling reference lists. In the former Soviet Union and Russia, references in most scientific journals are listed first in the order of the Russian alphabet and then in the order of the English alphabet.

This tradition reflects the closed character of Soviet/Russian science. In the Soviet Union, Soviet results had to be cited first. Scientists could not cite more foreign publications than publications of Soviet origin, especially before the 1980s.

Using Medline data, I found that in 1995, of a total of 163,007 publications related to human subjects, 1,784 were in Russian. If the hypothesis of influence of author surname on citations is to be taken seriously, culturally different subsets of the publication data need to be described separately.

I agree with Christopher Stubbs (*Nature* **388**, 320; 1997) that alphabetical authorship must be encouraged.

Some grant organizations ask for citations in applications only of publications where the applicant is the first author or co-author. At the same time, the first author is often the one who had the least input.

Vasily Vlassov

Saratov Medical University,
Box 1528,
Saratov 410601,
Russia
e-mail: vvvla@ssu.runnet.ru

Headless tadpoles and an informed public

Sir — In a recent leading article you mention work by my colleagues and me on the “headless frog embryos” (*Nature* **389**, 767; 1997). I should like to correct some misapprehensions.

First, we are at the University of Bath and not the University of Bristol.

Second, we did not release this as a news story to “a British television company”. The media interest arose from an article in the *Sunday Times*. I had spoken to the journalist involved on the previous Thursday thinking that he was simply writing a feature about the BBC Television *Horizon* programme in which our work was mentioned.

But, presumably by the decision of a subeditor, the article appeared not buried in the television section, but on the front page. By Sunday afternoon, everyone all over the world seemed to want to talk to us. In the various interviews, I have tried to explain the type of work we are doing in terms suitable for a lay audience. I have found the journalists I have dealt with generally serious and responsible, although they do not necessarily have control over later editing, headlines or interpretation.

People familiar with this field will doubtless be puzzled because they will know that headless tadpoles have been created many times before, although by chemical or surgical means rather than by the manipulation of genes.

But in the course of the past week I have learned that even specialist science journalists do not realize how much progress has taken place in developmental biology in the past 10 years. They are fascinated to learn that dramatic changes can be made in the anatomy of organisms by introducing or inhibiting one or two genes. They also usually do not realize how similar are the genetic mechanisms in different types of vertebrate animal, something that surprised the research community a few years ago.

This has been a useful opportunity to communicate some of the findings of developmental biology to a wider audience than usual. I agree with your leading article about the desirability of having an informed public. There are certain areas of biology where regulation of future research is inevitable, and the better informed the public becomes, the more likely it is that controls will be reasonable rather than restrictive.

Jonathan Slack

Developmental Biology Programme,
Department of Biology and Biochemistry,
University of Bath, Bath BA2 7AY, UK.
e-mail j.m.slack@bath.ac.uk

Mercury mining: profit or loss?

Sir — During the 1960s, the price of mercury reached its zenith — US\$560 per flask (1 flask = 34.47 kg). But the subsequent discovery of an increasingly large number of environmental and health problems caused by mercury has brought about a steady drop in prices and hundreds of publications about mercury pollution and anti-mercury social reactions¹. This has resulted in a serious financial crisis for mercury mines, the biggest of which are in Europe.

The mining district in Almaden, Spain, which produces 43,500 flasks a year (25 per cent of world production), is the largest reserve of mercury in the world (44 per cent). The Almaden mercury mines are the oldest (2,000 years) continuously active mines in the world. They played an important role during the colonial period because mercury was used in America in the amalgam method to extract silver and gold.

Two mining museums have recently been built in Almaden where visitors can see Roman tools, Arab amphorae for cooking cinnabar, interior-mine constructions, ancient furnaces and so on.

These interesting tourist attractions are, however, closed to the general public so that mercury mining and metallurgical extraction can proceed normally.

Not long ago, Mayasa, the company that runs the Almaden mines, announced a new plan to cope with its financial crisis; Mayasa now has 526 workers, compared with 2,000 in 1972, and it lost US\$13 million in 1996. The plan proposes a 34 per cent reduction in costs, dismissing 90 people, selling properties, diversifying its activities (geology consulting, boreholes, dairy and cattle products), prospecting for new mercury ore deposits in 40 mining areas and increasing mercury sales.

The European Union (EU) has recognized that Almaden, with 20,000 inhabitants, is a depressed rural area — deforested after 2,000 years of mercury furnaces and mine constructions. But, as the EU and the Spanish authorities turn a blind eye to Almaden's troubles and to the problem of global mercury pollution, Mayasa faces a critical financial situation alone, trying to protect its jobs and products.

Since 1982, Mayasa has developed a diversification programme to obtain additional cash from new sources. Of a yearly turnover of US\$33 million, only US\$5 million is now accounted for by the company's mercury mining activities.

(Mayasa sells 30,000 flasks a year and stores 13,500.)

Clearly, if the total workforce of 526 Mayasa workers produces a turnover of US\$33 million, and if only US\$5 million of this is produced by mercury sales, only 80 jobs ($\text{US\$5 million} \times 526 \text{ workers} / \text{US\$33 million}$) are involved in the process of extracting, refining and exporting mercury. So 446 jobs should be accounted for by new alternative activities of the 1982 plan.

The present distribution of global mercury use² has been extensively criticized in scientific journals^{3,4}. There are viable substitutes for almost all these uses: diaphragm and membrane cells in the electrolytic production of chlorine and caustic soda, gold recovery by cyanidation, and lithium, zinc-air or nickel-cadmium cells in batteries^{5,6}.

If the EU wanted to reduce global mercury pollution, it would need to discriminate between mercury buyers according to final use. We propose that the EU should impose controls on the European production of mercury (Spain, Italy, Slovenia), inviting the United States (New Idria, New Almaden) to join this moratorium.

The EU must realize that helping impoverished mercury mining regions of Europe by, for example, providing financial aid for tourism in Almaden would bring

about a huge reduction in global mercury pollution.

Javier Garcia-Guinea

Matthew Harffy

Museo Nacional de Ciencias Naturales,

C/ Jose Gutierrez Abascal 2,

28006 Madrid, Spain

e-mail: guinea@fresno.csic.es

1. Wheatley, B. & Wyzga, R. *Special Issue of Water, Air & Soil Pollution* **97**, 1–2 (1997).
2. Hernandez Sobrino, A. *Las Minas de Almadén* (Minas de Almadén y Arrayanes, Madrid, 1995).
3. Ellingsen, D. G. *et al. Br. J. Ind. Med.* **50** (10), 881–887 (1993).
4. Nriagu, J. O. *et al. Nature* **356**, 389 (1992).
5. Reese, R. G. *Mineral commodity summaries: Mercury* 102–103. (US Dept of the Interior, Bureau of Mines and US Geological Survey, Washington DC, 1991).
6. Dunn, J. G., Ibrado, A. S. & Graham, J. *Miner. Engineering* **8** (4–5), 459–471 (1995).

Getting rid of mosquitoes

Sir—I read with interest the News report “Consortium aims to revive sterile-mosquito project” (*Nature* **389**, 6; 1997).

I was a member of the Advisory Group established by the World Health Organization (WHO) to plan, coordinate and oversee feasibility studies on genetic manipulation of three major mosquito vectors of human disease (*Aedes aegypti*, *Anopheles stephensi* and *Culex fatigans*) in Delhi, India, during the early 1970s.

A good deal of useful data were collected

on the field biology and ecology of the target species and on small-scale field releases of appropriate genetic mechanisms including sterile-insect technique (SIT). But the Delhi project was terminated prematurely for political reasons. Consequently, the concept of the use of genetic control of mosquitoes was not adequately evaluated. Until proved otherwise, the use of classical genetic methods including SIT cannot and should not be written off. I therefore applaud the latest effort to revisit and possibly to resurrect this field. I should like, however, to make the following comments:

(1) The suggestion that spending \$9 million “just once [would] eliminate malaria forever using SIT” is erroneous. SIT control of “urban island” populations of *A. stephensi* in South India may well be possible, but it will certainly not be a one-off ‘magic bullet’.

(2) One major lesson learned from WHO-sponsored field releases in India was that mosquito immigration from surrounding areas was an important factor. Thus the statement attributed to Christopher Curtis that “If we get rid of *A. stephensi* in towns using SIT, no replacement is possible and the area can remain free of malaria forever” may be unrealistic.

(3) Female monogamy is not a requirement for successful application of SIT.

(4) The price of \$12 million “needed to start the project” may be over-ambitious, particularly in view of the “Multinational Initiative in Malaria” being established by the world’s foremost funding agencies.

Nevertheless, in view of the scientific infrastructure, the facilities, a sufficiently trained workforce and the high incidence of malaria, India still provides perhaps the best place to test the premise of genetic control for disease prevention.

Karamjit S. Rai

Department of Biological Sciences,

University of Notre Dame,

Notre Dame, Indiana 46556, USA

e-mail: karamjit.s.raij@nd.edu

Funding fault

Sir—Sir Hermann Bondi refers to the waste of time when there are too many claimants for available funding (*Nature* **388**, 709; 1997). If ten claimants each apply to the same ten sources of funding and they each score one success and nine failures, this shows not that there are too many scientists but that the funding system is in need of reform.

Henry Haslam

British Geological Survey,

Keyworth,

Nottingham NG12 5GG, UK

Galaxies take the distance record

Kenneth Lanzetta

A pair of forming galaxies are now the most distant known objects. Their knotty, irregular structure – visible with the help of an intervening gravitational lens – gives a close-up view of star formation in the early Universe.

There was a time not long ago when the most distant objects known by far were quasars — exceptional objects powered by hot gas falling onto super-massive black holes. This is easy enough to understand: quasars are the most luminous objects that exist, outshining ordinary galaxies by large factors. Because they are so luminous, quasars appear relatively bright even when viewed across the vast stretches of the Universe.

But over the past two years or so, extremely sensitive observations with the Hubble Space Telescope and the Keck telescope have revealed a Universe of very distant galaxies — unexceptional objects which will by now have become like the ordinary elliptical, spiral and irregular aggregates of stars, gas and dust that we know nearby. These ordinary galaxies are far less luminous than quasars, so they are more difficult to see. Nevertheless, Marijn Franx and collaborators have now reported¹ the spectroscopic identification of a pair of galaxies at a redshift of $z = 4.92$. For the first time since the discovery of quasars in 1963, the most distant objects with spectroscopic identifications are ordinary galaxies rather than quasars.

Setting a new distance record might seem impressive enough, but it is not really distance that makes these galaxies of interest to cosmologists. In cosmology it is time that is paramount, and distance and time are inextricably linked. Because light travels at a finite speed, objects are seen as they were in

the past. In ordinary experience, the time delay between an event and our perception of the event is of little consequence, but in cosmology — where the distances are enormous — the delay can be many billions of years.

Cosmologists express distance, or equivalently 'look-back' time, by the redshift z , which is a measure of the fractional wavelength shift suffered by light as a result of the expansion of the Universe. At a redshift of $z = 4.92$, the galaxies identified by Franx and collaborators are seen as they were when the Universe was 20% of its current size, and only about 10% of its current age — or less than a couple of billion years after the Big Bang, given plausible estimates of the age of the Universe. That means that these two galaxies must be at a very early stage of development.

These galaxies are of particular interest, not just because of their great distances, but also because their images have been magnified by the gravitational field of an intervening cluster of galaxies (Fig. 1). This 'gravitational lens' effect — a consequence of general relativity that was first worked out by Einstein in 1936 — makes the galaxies appear both brighter and larger than they otherwise would. Plausible models of the intervening cluster suggest a magnification factor of up to around ten. Distant galaxies are difficult to study precisely because they appear faint and small, so the added boost provided by the

intervening cluster allows the galaxies to be seen in unprecedented detail.

And just what do the new observations reveal? Compared with the intricate and beautiful patterns of nearby spiral galaxies, the galaxies identified by Franx and collaborators are not much to look at. The brighter of the two (G1; Fig. 2) is resolved into one bright knot and several fainter knots, whereas the fainter (G2) is resolved into just a single knot. The knots are compact regions of intense star formation, similar to those seen in nearby 'starburst' galaxies. But the knots in galaxies G1 and G2 are even more luminous than the knots typical of nearby starburst galaxies, indicating higher star formation rates in the distant past.

This is not at all unexpected. Many models predict that a substantial fraction of the stars in today's galaxies were formed at early times, when large quantities of gas could have provided the raw material. At the rate at which galaxy G1 is forming stars, a large portion of the central bulge of our own Milky Way galaxy could be formed in just 100 million years.

But the galaxies show evidence of gas outflows, which are presumably driven by the intense star formation activity. If outflows rapidly disperse the gas from which the stars are formed, then the star formation episodes must be brief. This might explain the knotty, irregular structure of the galaxies. Galaxies of redshift z near 3 appear less irregular², but these lower-redshift galaxies are not magnified by gravitational lenses and so are seen in much less detail.

Interpreting galaxies G1 and G2 is complicated by the same factor that complicates the interpretation of other distant galaxies: dust. Dust obscures light, especially the ultraviolet and blue light emitted by young stars. (This is the reason that the setting sun appears red. Dust in Earth's atmosphere scatters and absorbs blue light but leaves red light relatively unaffected.) If galaxies G1 and G2 are heavily obscured by dust, then their star

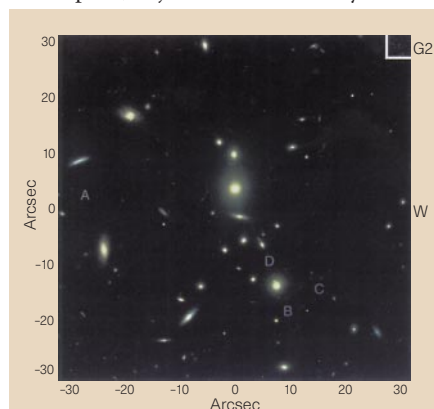


Figure 1 This cluster of galaxies acts as a gravitational lens, magnifying much more distant galaxies behind it.

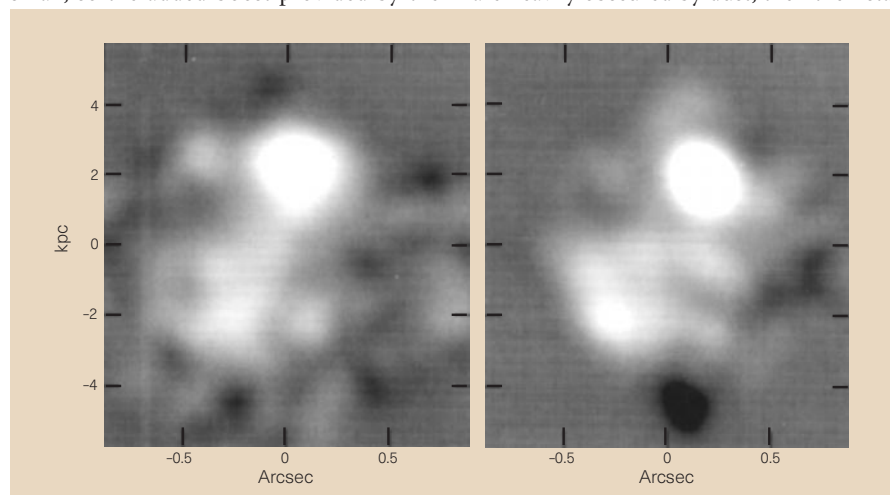


Figure 2 Two images of the galaxy G1, each reconstructed from a separate gravitationally lensed arc. At a redshift of 4.92, it and its partner G2 are the most distant known galaxies.

formation rates could be much higher than those directly indicated by the observations.

Just how much of the ultraviolet light of distant galaxies is obscured by dust is a question that profoundly affects our understanding of how galaxies form and evolve. Some argue for a high degree of obscuration, as in nearby dusty starburst galaxies^{3,4}. If that is the case, galaxy formation was early and rapid. Others argue for low obscuration, reasoning that if significant early star formation were hidden from view, then there should be more old stars in today's galaxies than are actually seen⁵. If that is so, galaxy formation was late and slow. The difference is important because it amounts to the difference between galaxy formation by 'monolithic collapse' of protogalactic gas clouds and formation by hierarchical merging of subgalactic fragments.

Deciding between these possibilities is not going to be easy. One strategy is to observe distant galaxies at infrared wavelengths, where the effects of dust obscuration are less severe. This capability is just within reach of large ground-based telescopes today but should become commonplace as adaptive optics techniques — which compensate for the blurring effect of Earth's atmosphere — are developed and implemented. If the pace of progress continues as it has over the past few years, an answer is likely to be found soon. □

1. Franx, M., Illingworth, G. D., Kelson, D. D., van Dokkum, P. G. & Tran, K.-V. *Astrophys. J.* **486**, L75–L78 (1997).
2. Giallisco, M., Steidel, C. C. & Macchetto, D. *Astrophys. J.* **470**, 189–194 (1996).
3. Meurer, G. R., Heckman, T. M., Lehnert, M. D., Leitherer, C. & Lowenthal, J. *Astron. J.* (in the press).
4. Rowan-Robinson, M. et al. *Mon. Not. R. Astron. Soc.* **289**, 490–496 (1997).
5. Madau, P. preprint astro-ph/9707141 on xxx.lanl.gov

Apoptosis

A Bad kinase makes good

Thomas F. Franke and Lewis C. Cantley

The regulation of cell survival is a central feature in animal development and in diseases such as cancer. Results published over the past year implicate phosphoinositide 3-kinase (PI3K) and its downstream target, the serine/threonine protein kinase Akt, in a pathway that conveys survival signals from various cell-surface receptors¹. Two papers in *Science*² and *Cell*³ now report a missing piece in this signalling pathway. They show that Akt phosphorylates a critical serine residue on Bad, a protein that promotes cell death (apoptosis) by binding to and blocking the activity of Bcl-x_L, a cell

survival factor. Upon phosphorylation, Bad dissociates from Bcl-x_L, which is then free to resume its activity as a suppressor of cell death (Fig. 1).

The Bcl-2 family of proteins has been implicated in cell survival decisions. Bcl-2 and Bcl-x_L suppress apoptosis, in part by blocking the release of cytochrome *c* from mitochondria, a critical step in the activation of the caspase protease cascade. Other family members, such as Bad and Bax, interfere with the activity of Bcl-2 or Bcl-x_L by binding to them and forming non-functional heterodimers⁴. In response to interleukin-3, for

example, Bad is phosphorylated at two sites, dissociates from Bcl-x_L and then interacts with the phosphoserine-binding protein, 14-3-3 (ref. 5). The released Bcl-x_L then promotes cell survival by blocking the caspase protease cascade.

Del Peso *et al.*² now show that PI3K activity is necessary for the phosphorylation of Bad after interleukin-3 treatment of cells. Moreover, they have identified Akt as the kinase that phosphorylates Bad on a site that is required for its interaction with 14-3-3. Datta and colleagues³ independently identified Bad as a substrate for the Akt kinase and showed that of two potential phosphorylation sites in mouse Bad, only one (Ser 136) is phosphorylated by Akt. Furthermore, they demonstrate that Akt overexpression prevents Bad-induced cell death of cerebellar granule neurons, suggesting that Akt-dependent cell survival is mediated in part by phosphorylation of Bad.

Other studies further clarify how Akt itself is regulated. Previous research has shown that Akt directly binds to phosphoinositide products of PI3K by an N-terminal pleckstrin homology (PH) domain¹. Although this interaction causes partial activation of Akt, full activation requires that Akt be phosphorylated by other serine/threonine protein kinases. One such kinase, phosphoinositide-dependent kinase-1 (PDK1), which phosphorylates and activates Akt *in vitro*, has now been purified and cloned^{6–8}. PDK1 has a carboxy-terminal PH domain that apparently is involved in its activation by phosphoinositide products of PI3K (Fig. 1). Taken together, these findings identify a sequence of signalling events that originates with the binding of survival factors to cell-surface receptors and results in the generation of second messenger molecules through the activation of PI3K. These second-messengers induce the activity of the Akt kinase by binding to the PH domains of Akt and PDK1, resulting in phosphorylation of residues on Akt, which are critical for the regulation of its activity. Activated Akt then phosphorylates the pro-apoptotic protein Bad. Phosphorylated Bad is translocated from the mitochondria to the cytosol through its interaction with 14-3-3, and the function of Bcl-x_L, the gatekeeper of cytochrome *c* efflux at the mitochondrial membrane, is restored. By blocking the release of cytochrome *c* from the mitochondria, caspase activation and further execution of the cell-death cascade is prevented.

The discovery that Akt is a Bad kinase provides an intriguing explanation for the role of PI3K and Akt in cell survival. However, not all survival signals require PI3K activity and Akt-independent survival pathways exist⁹. Furthermore, it is likely that there are additional targets of Akt in cell survival because Bad is not ubiquitously expressed. Moreover, PI3K and Akt feature in other signalling pathways

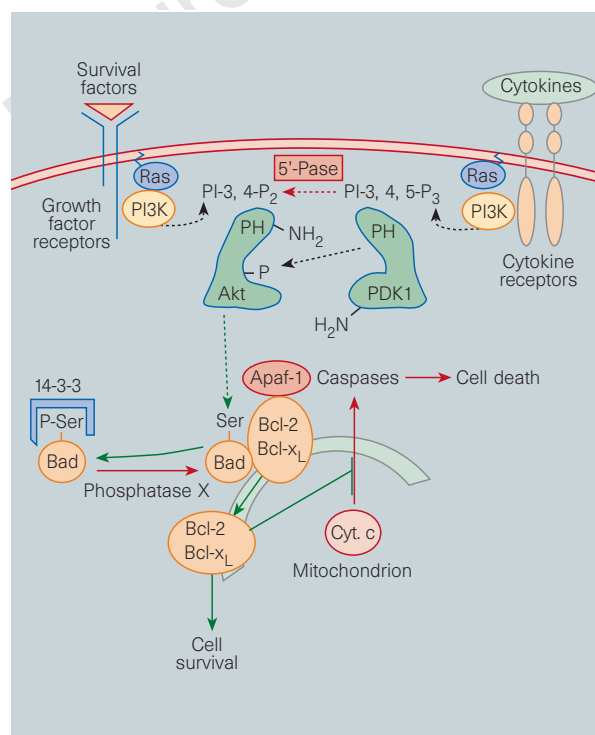


Figure 1 The PI3K/Akt signalling cascade and cell survival. When various extracellular signals including growth factors and cytokines bind to their cell-surface receptors, intracellular phosphoinositide 3-kinase (PI3K) becomes activated. This results in the generation of two lipid products (PI-3,4-P₂ and PI-3,4,5-P₃; see page 123 of this issue), which act as second-messenger molecules and activate the serine/threonine kinases Akt and PDK1 (through their PH domains). Activated Akt phosphorylates the pro-apoptotic factor Bad on a serine residue, resulting in its dissociation from Bcl-x_L and its association with 14-3-3. Released Bcl-x_L is then capable of suppressing cell-death pathways that involve the activity of Apaf-1, cytochrome *c* (Cyt. *c*) and the caspase protease cascade¹⁰.

that are apparently unrelated to cell survival, such as those important in cellular responses to insulin. Phosphoinositide products of PI3K also mediate responses unrelated to activation of Akt. So downstream branch points exist at each step of the signalling cascade and each step has multiple upstream regulators. For example, there are mechanisms of Akt activation that do not require PI3K (ref. 1) and there is a pathway involving phosphorylation (at Ser 112) of Bad that is independent of Akt (ref. 3). Thus, the complete signalling cascade revealed by these reports is but a skeleton and more research is needed to put meat on the bones. □

Thomas F. Franke is in the Department of Pharmacology, College of Physicians & Surgeons of Columbia University, 630 West 168th Street, New

York, New York 10032, USA, and in the Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec H3A 2T5, Canada. Lewis C. Cantley is in the Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

1. Franke, T. F., Kaplan, D. R. & Cantley, L. C. *Cell* **88**, 435–437 (1997).
2. del Peso, L., González-García, M., Page, C., Herrera, R. & Nuñez, G. *Science* **278**, 687–689 (1997).
3. Datta, S. R. *et al.* *Cell* **91**, 231–241 (1997).
4. Golstein, P. *Science* **275**, 1081–1082 (1997).
5. Zha, J., Harada, H., Yang, E., Jockel, J. & Korsmeyer, S. J. *Cell* **87**, 619–628 (1996).
6. Alessi, D. R. *et al.* *Curr. Biol.* **7**, 261–269 (1997).
7. Stokoe, D. *et al.* *Science* **277**, 567–570 (1997).
8. Alessi, D. R. *et al.* *Curr. Biol.* **7**, 776–789 (1997).
9. Philpott, K. L., McCarthy, M. J., Klippel, A. & Rubin, L. L. *J. Cell Biol.* **139**, 809–815 (1997).
10. Hengartner, M. O. *Nature* **388**, 714–715 (1997).

Sudden end of an interglacial

Scott Lehman

Much of humankind's development as an agrarian and then industrial society has occurred during the current Holocene interglacial period, a roughly 9,000-year stretch of anomalously mild and stable climate (although it may not seem so at times). In geological time, such conditions are a rarity, and they can by no means be taken for granted. Hence the interest in finding out how stable interglacials really are, and what processes might act to bring them to a sudden close.

On page 154 of this issue¹ Adkins and colleagues present one of the most detailed pictures yet of ocean conditions during the previous interglacial. Using a new geochemical approach to sediment dating (see box, overleaf), they first show that the interval during which global ice volume was comparable to or less than today's lasted from about 129,000 to 119,000 years ago. They then show that the 'conveyor belt' circulation which today carries ocean heat north from the tropics and warms much of Europe remained strong and relatively steady throughout the interval. Most interestingly, they also find evidence that the onset of ice growth marking the end of the interglacial was accompanied by a sudden reduction in conveyor circulation, which took only 400 years and from which the climate system apparently never recovered.

In Europe, detailed pollen spectra from the last interglacial have long been available. Most were originally interpreted to indicate an early interval of slightly higher-than-modern temperatures lasting a few millennia, followed by gradual cooling of a few degrees centigrade. This view was upset briefly in 1993 when the first oxygen-isotope results from presumed last-interglacial ice

near the base of one of two new Greenland ice cores showed evidence of sudden 5–10 °C changes in temperature. These results later turned out to be artefacts of ice-flow disturbance². Nonetheless, they led to renewed study of the European sites, and to revised interpretations consistent with climate volatility (some of which have since proved questionable³). By concentrating on the oceans, Adkins and colleagues side-step the complex relationship between vegetation and climate. They also build on work showing that uncontested signs of abrupt temperature change during glacial times (some 75,000–10,000 years ago) are typically accompanied, and probably caused, by large alterations in ocean circulation^{4–6}.

The new results come from a 53-m-long sediment core from the Bermuda Rise, which is located 4.5 km beneath the surface of the Sargasso Sea in the western North Atlantic. The site is unusual in that deep gyres, driven by the same forces that propel the Gulf

Stream near the surface, act to re-suspend fine-grained sediments over a broad region and deposit them preferentially on the Bermuda Rise at rates 10–100 times the average for the open ocean. In consequence, the core provides exceptional resolution in time.

The site also lies in the vertical mixing zone between nutrient-rich deep waters that form around Antarctica and nutrient-poor North Atlantic Deep Water (NADW) that constitutes the lower limb of the Atlantic's heat-carrying conveyor circulation. Geochemical proxies of past nutrient levels at the Bermuda Rise have already provided some of the clearest records of deep circulation change – in fact, they have been crucial in establishing the relationship between past reductions in conveyor strength and evidence of cooling in the North Atlantic region^{6,7}. Until recently, the longest cores from the area measured 25 m and sampled only the past 80,000 years of sedimentation. The new cores punch back to 160,000 years, and for this we owe a great deal to the equipment and crew of the French research vessel *Marion Dufresne*, and to international support for its mission.

Conventional measurements of the ratio of ¹⁸O to ¹⁶O in the fossil shells of bottom-dwelling foraminifera provide the stratigraphic backbone of Adkins and colleagues' results. The ratio increases in sea water (and in carbonate shells formed in isotopic equilibrium with sea water) during periods of global ice-sheet growth as highly fractionated precipitation enriched in H₂¹⁶O is preferentially locked away in glacier ice. The periodic growth and decay of continental ice sheets has thus produced a characteristic suite of oxygen isotope changes, and these have been matched to variations in the distribution of solar radiation whose timing is precisely determined from orbital mechanics.

This 'orbital tuning' places the beginning of the isotopic transition coming into the last interglacial at about 131,000 years ago, and



Figure 1 Core vessel. The oceanographic ship *Marion Dufresne*, from which the 53-m core described by Adkins *et al.*¹ was drilled. The core and the data gleaned from it constitute the first big haul for the International Marine Global Change Study.

IRPT/M. LE COZ

the transition into the subsequent glacial period at roughly 114,000 years. In past studies it has had to be assumed that deposition between these 'mile posts' has been constant, which may be so at typical open-ocean locations where deposition rates are slow. But Adkins *et al.* looked at the Bermuda Rise precisely because rates of deposition there are much greater, and undoubtedly much more variable.

To produce a timescale that accounts for such changes, they take advantage of the relative constancy in the rate of deposition of the isotope ^{230}Th in the deep sea and its unusual concentration by redeposition at the Bermuda Rise. This is a real innovation, which allows much more precise dating between the mile posts (see box). For example, the resulting, refined timescale suggests that the period during which global ice volume was comparable to or less than today's lasted from 129,000 to 119,000 years ago, or for about 10,000 years. This is 3,000 years less than implied by simple linear interpolation between mile posts. The duration of transitional events is also well constrained for the first time.

What was the ocean doing during the last interglacial? Adkins and colleagues' measurements of the cadmium/calcium ratio in benthic foraminifera (a proxy for seawater nutrient concentration) show that the ratio remained much the same as during the Holocene⁶, implying that the proportion of NADW at the Bermuda Rise was similar to today's throughout the period of minimum ice volume. ^{230}Th -derived fluxes of carbonate and fine-grained detritus also remained nearly constant. These observations seem to rule out the possibility of any large swings in circulation, surface productivity and land-sea sediment transfer lasting more than a few hundred years.

The largest geochemical change in the new record is seen during the ice-sheet melting phase leading into the interglacial, when Cd/Ca ratios shot up to values usually encountered in the nutrient-rich Pacific. Similar evidence from elsewhere in the Atlantic has been taken to suggest that fresh water from rapidly melting ice sheets swamped the conveyor, stemming the flow of nutrient-poor NADW^{8,9}. Adkins and co-authors' evidence for a large increase in the transfer of sediment

from land to sea early in the high Cd/Ca event is strong support for this view.

What has not been seen before is the evidence from the Cd/Ca ratio for a reduction in NADW flow at the end of the interglacial, just at the point where benthic $^{18}\text{O}/^{16}\text{O}$ ratios begin to rise. This Cd/Ca event is not large compared with the earlier one, but its significance is confirmed by an increase in carbonate dissolution, as expected if NADW gave way to a nutrient-rich, and therefore CO_2 -rich, water mass. The detail is unprecedented and so one cannot be entirely sure that the initial oxygen-isotope changes are the first signs of irreversible ice-sheet growth (differences in temperature and $^{18}\text{O}/^{16}\text{O}$ ratio between water masses may factor in). But this is the simplest explanation.

If true, what caused the about-face in the climate system? There is no simple answer. But (tantalizingly) many characteristics of the terminal interglacial event recorded by Adkins *et al.* are similar to those reported¹⁰ from the Bermuda Rise for the Little Ice Age, a 'cold snap' which occurred a few hundred years ago. Arctic field geologists have long suspected that the end of the present interglacial was then very nearly at hand, as semi-permanent snowfields threatened to coalesce over the high plateaux of Labrador and Baffin Island¹¹. If the coalescence had been complete, albedo feedback along with decreasing summer insolation (on the decline locally for the past 9,000 years) might have triumphed over interglacial warmth. Numerical models suggest that such a chain of events can be unleashed with very little forcing¹².

In this context, a slight weakening of the ocean circulation can be likened to a sledgehammer blow — one which could conceivably bring the present interglacial to its knees. Such a possibility only adds to concerns about the potential impact of increased greenhouse forcing on the oceans. □

Scott Lehman is in the Department of Geological Sciences and the Institute of Arctic and Alpine Research, Campus Box 450, University of Colorado, Boulder, Colorado 80309, USA.

e-mail: lehman@spot.colorado.edu

- Adkins, J. F., Boyle, E. A., Keigwin, L. & Cortijo, E. *Nature* **390**, 154–156 (1997).
- Alley, R. B. *et al.* *Nature* **373**, 393–394 (1995).
- Auby, B. & Tauber, H. *Nature* **376**, 27–28 (1995).
- Broecker, W. S., Peteet, D. M. & Rind, D. *Nature* **315**, 21–26 (1985).
- Bond, G. *et al.* *Nature* **365**, 143–147 (1993).
- Boyle, E. A. & Keigwin, L. *Nature* **330**, 35–40 (1987).
- Keigwin, L. D. & Jones, G. A. *J. Geophys. Res.* **99**, 12397–12410 (1994).
- Sarnthein, M. & Tiedemann, R. *Paleoceanography* **5**, 1041–1055 (1990).
- Oppo, D., Horowitz, M. & Lehman, S. J. *Paleoceanography* **12**, 51–63 (1997).
- Keigwin, L. D. *Science* **274**, 1504–1508 (1996).
- Andrews, J. T. & Barry, R. *Annu. Rev. Earth Planet. Sci.* **6**, 205–228 (1978).
- North, G. R. *J. Atmos. Sci.* **41**, 3390–3395 (1984).
- Suman, D. O. & Bacon, M. P. *Deep Sea Res.* **36**, 869–878 (1989).
- Adkins, J. F., Boyle, E. A. & Keigwin, L. D. *EOS Trans. AGU* **76**, F282 (1995).

Accordioning with thorium

In the work discussed here, Adkins *et al.*¹ demonstrate that measurements of ^{230}Th , a uranium decay product, can be used to account for changes in sediment deposition rate between the orbitally derived 'mile posts' for the beginning and end of the last interglacial period.

Uranium is soluble and of nearly constant concentration in sea water, whereas thorium adsorbs rapidly to sedimenting detrital particles. Decay of ^{234}U to ^{230}Th within sea water therefore produces a flux of ^{230}Th to the sea floor that depends only on water depth. After accounting for some production of ^{230}Th in sediments (from sedimentary U, which is easily measured) and some loss due to decay (^{230}Th is itself radioactive), one can back-calculate the amount of ^{230}Th at the

time of deposition, which however requires independent estimates of age. These are, of course, lacking between the mile posts. But precise Th and U measurements in radiocarbon-dated Holocene sediments of the Bermuda Rise^{13,14} have shown that age- and U-corrected ^{230}Th values were often considerably higher than expected solely from local seawater production, indicating that regionally deposited ^{230}Th had been swept up by bottom currents and preferentially redeposited at the Bermuda Rise.

The amount of this sediment 'focusing' can be calculated simply as the excess ^{230}Th measured and corrected for decay since the time of deposition divided by that expected from production in overlying sea water in a given increment of time¹⁵. At the Bermuda Rise this

'focusing factor' (F) varies inversely with the percentage by weight of biogenic carbonate, presumably because the same climatic forces that promote delivery of terrestrial detritus to the deep western Atlantic and, consequently, dilution of the near-constant surface production of carbonate, also act to invigorate recirculation at depth.

Whatever the cause, Adkins *et al.* realized they could estimate F from carbonate proxy data, and then use their ^{230}Th results to calculate the time elapsed between measurements — which in this case were made at astonishingly dense 2-cm intervals. Their technique is still heavily reliant on the accuracy of the orbitally derived mile posts, but this way of refining the timescale will be very useful even if those posts are eventually moved. **S.L.**

Human evolution

Ecce Homo — behold mankind

Bernard Wood

We are incorrigibly curious about our origins, yet we are curiously ambivalent when confronted with the prospect of obtaining sound evidence about the emergence of modern humans 250,000 to 150,000 years ago. Developments in molecular biology have made available a new generation of molecular markers that can record events in the evolution of *Homo sapiens* that cannot be detected with conventional methods. In combination with advances that enable whole genotypes to be rapidly screened for distinctive sequences, they promise to provide a much more precise picture of the relationships between regional populations of modern humans. Last month, the Cold Spring Harbor Laboratory hosted a meeting* to take stock of the progress that has been made towards reconstructing our evolutionary history over the past 250,000 years.

The first attempts to recover what Darwin called the 'perfect pedigree of mankind' were based on head shape. However, once the genetic basis of inheritance was understood, and advances in biochemistry enabled researchers to detect variation at the molecular level, attention turned to blood groups, then to plasma proteins and later to enzymes, in the pursuit of what became known as molecular anthropology¹. The deciphering of the genetic code eliminated the need to rely on phenotypic proxies for information about propinquity, and molecular anthropology was upstaged by direct comparisons between the genomes of individuals sampled from regional populations of modern humans.

The first rigorous attempts to relate regional populations used gene frequencies². These are converted to genetic distances and then the relationships between populations are represented using a variety of 'tree-building' techniques. The variations in gene frequencies of 'trees' that are 'rooted' can be represented as 'surfaces' on maps³. When several of these apparently independent genetic systems have their root in the same geographical location this is interpreted as the likely origin of that population⁴. But frequency data are not readily resolved into the type of 'primitive' and 'derived' states that provide reliable evidence about the relationships between populations.

Although this first generation of studies could call upon substantial allelic variation for some of the genetic markers — HLA-A and HLA-B histocompatibility complex antigens, for example — few, if any, of the alleles are exclusive to any of the regional

groups. If less widely distributed markers could be found, then these would define the evolutionary relationships between regional populations of modern humans with much greater precision. Such markers would be particularly effective if their own evolutionary history could be traced. Three polymorphic systems with the potential to provide such information are Alu inserts; short tandem repeats (microsatellites); and single-nucleotide polymorphisms.

A substantial part of our genome consists of short interspersed repeated DNA sequences about 300 base pairs in length. These are recognized and cleaved by the restriction enzyme, *AluI*, and so are known as the Alu family of sequences. They are probably derived from an ancestral RNA gene that has been progressively amplified by a process called retroposition, and are only found in the genomes of primates. Each human genome contains upwards of half-a-million of these sequences, which make up some 5% of our DNA. Alu sequences have several advantages over traditional genetic markers. They are easily screened and stable (that is, newly inserted Alu elements rarely undergo

deletion) and they represent unique events. Moreover, it is known that the Alu element is absent in the ancestral state⁵.

At the meeting, M. Stoneking (Penn. State Univ.) reported the results of an analysis of eight Alu loci in the nuclear DNA of 1,500 people drawn from populations around the world. From their results, the investigators estimated that the major split between African and non-African populations occurred about 140,000 years ago; this date is consistent with evidence from mitochondrial DNA (mtDNA). Notably, Stoneking found that population samples from Sahul — that is, from Australia and New Guinea — are as close to the presumed ancestral state of the human Alu family as samples from African populations. This suggests that a migration to Southeast Asia may have been one of the first 'out of Africa' excursions of modern humans.

Whereas mtDNA traces phylogeny through the maternal line, evidence from the Y chromosome descends through the male line. The different relationships of the two systems provide the potential to investigate whether one, or other, sex migrated from the founding populations of modern humans. So far, attention has centred on the non-recombining part of the Y chromosome (P. Underhill, Stanford; M. Hammer, Univ. of Arizona). Underhill reported success in separating duplexes between the non-recombin-

Evolutionary biology

Fungal foray

Some of the homobasidiomycete group of fungi, which includes most of the mushrooms with which we are familiar, have earned exotic names such as bird's nest and puffball because of the appearance of their fruiting bodies (the part of the fungus that produces the spores).

Traditional classification divides this group into the gilled fungi (which have fruiting bodies with a cap and a gilled underpart containing the spores) and the puffballs and relatives (in which the spores are enclosed within variously shaped fruiting bodies). The gilled species expel their spores by an elaborate ballistic mechanism (ballistospory) whereas the puffballs for instance slowly crumble, gently releasing their cargo. Although fruiting bodies are diverse in appearance their anatomy is very simple. This, coupled with the dearth of fossil evidence, has made it difficult to unravel the fungal family tree.

Writing in *The Proceedings of the National Academy of Sciences* (94, 12002–12006; 1997), David Hibbett and colleagues have tackled the complexities of fungal evolution by molecular analyses of



81 species. They compared the nucleotide sequences of various genes and drew up a family tree based on the similarity of sequences between different species.

It turns out that the same gill-and-cap style of fruiting body evolved at least six times. Furthermore, some of the puffballs and their relatives arose from their gilled cousins whereas others, such as the bird's nest fungus (*Crucibulum laeve*; shown here), developed independently. It seems that the group including the puffballs never evolved into gilled mushrooms — the authors surmise that ballistospory can develop into other forms of dispersal, but the reverse has not occurred.

Orla Smith

*Human Evolution Meeting, Cold Spring Harbor Laboratory, New York, USA, 4–8 October 1997.

ing regions of Y chromosomes using denaturing high-performance liquid chromatography with alkylated particles to exploit the differential retention of homo- and heteroduplex DNA molecules⁶. Although the report presented at the meeting was based on 93 biallelic markers, this group has detected more than 100 novel polymorphisms to date. They screened more than 900 Y chromosomes drawn from populations worldwide and identified 71 haplotypes. These could be broken down into five major haplogroups, upon which substantial regional localization was superimposed.

An ancient A-to-T transversion event determined the earliest branch in the tree, with the ancestral A allele being limited to a subset of modern Africans. All other contemporary peoples across the globe, together with the majority of African populations, display the mutant T allele. This suggests that most modern Y chromosomes are inherited from a single African male individual, a finding that is mirrored in females, as shown by evidence from mtDNA. But, in another presentation, P. de Knijff (Leiden Univ.) suggested that the relatively high recurrent mutation rate limited the effectiveness of Y microsatellite data to the interpretation of genetic affinities over the past 50,000 years. If de Knijff is correct, and if the origin of modern humans was 250,000 to 150,000 years ago, then this high rate would greatly reduce the usefulness of Y-chromosome polymorphisms for determining population relationships at this crucial time in human evolutionary history.

An alternative strategy, adopted by K. Weiss (Penn. State Univ.), is to opt for knowing a lot about a little. His group have concentrated on sequencing 10 kilobases of DNA in 142 alleles of a single gene, which encodes lipoprotein lipase. They found that even in their single gene system there is less 'order' and more recombination than is assumed by conventional models. Thus, the data from traditional and new genetic markers may not be as 'clean' as many researchers are assuming. A global study of mtDNA using restriction analysis and control region sequencing (D. Wallace, Emory Univ.) reaffirmed the relative antiquity of mtDNA samples from modern African populations. This study also points to some discrepancies between phenotypic and molecular evidence because the results show that samples of mtDNA from the Western and Eastern Pygmies of Central Africa are remarkably dissimilar, despite the close physical similarity of the two groups.

Where does all this new information leave us with respect to the origin of modern humans? Although none of the new evidence presented at the meeting contradicted the proposition that Africa is the source of much of modern human genetic variation, it failed to add more precision to the timing of the latest 'out of Africa' diaspora. Many factors, including population size and the degree to

which physical barriers affected genetic admixture, influence estimates of the timing of these major population migrations⁷. We must hope that the new, more precise, analytical methods that are being developed will eventually add much-needed accuracy and precision to these attempts to trace modern human evolutionary history. □

Bernard Wood is in the Department of Anthropology, George Washington University,

Washington DC 20052, USA.

e-mail: bwood@gwis2.circ.gwu.edu

1. Zuckerkandl, E. in *Classification and Human Evolution* (ed. Washburn, S.) 243–272 (Aldine, Chicago, 1963).
2. Cavalli-Sforza, L. L. *Proc. R. Soc. Lond.* **164**, 362–379 (1966).
3. Cavalli-Sforza, L. L. *et al. Cold Spring Harbor Symp. Quant. Biol.* **51**, 411–417 (1987).
4. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, 1995).
5. Batzer, M. *et al. J. Mol. Evol.* **42**, 22–29 (1996).
6. Underhill, P. A. *et al. Genome Res.* **10**, 996–1005 (1997).
7. Harpending, H. *Pers. Biol. Med.* **37**, 384–394 (1994).

Nuclear physics

Doubly magical gamma decays

Philip Woods

The harmony of the periodic table in chemistry is a manifestation of the underlying quantum shell structure of the atom. Noble-gas elements with full electron shells are unreactive and have high ionization energies. The chemical properties of other elements can be understood in relation to these closed shells, whose location is solely defined by the atomic number Z , the total charge of the atomic nucleus. Similarly, the nucleus can be modelled in terms of neutron and proton shells, closed-shell nuclei having 'magic numbers' of neutrons or protons. But the presence of two varieties of particle leads to more richness and complexity in nuclei than in atoms. This can produce surprises, such as the one that Górska *et al.*¹ have now found in cadmium-98, a relation of the doubly magic nucleus, tin-100.

The first evidence for nuclear shell structure came from measurements of the masses of stable isotopes, and their natural abundances. It was found that nuclei with certain neutron (N) and proton (Z) numbers are abundant, and have high binding energies—they are very stable, in other words. Modern experiments are able to probe shell structures of nuclei very far from stability. Particularly revealing information can be obtained in studies of gamma-ray emission from excited states of nuclei at or near shell closures. Such states should have simple configurations that can be readily interpreted in terms of excitations with respect to a stable

core. In general, one expects nuclei with closed neutron or proton shells to have high-lying excited states compared with the ground state, because particles must be promoted across a major shell gap to be excited.

This is indeed the case for the stable isotope ^{40}Ca , which has magic numbers of both neutrons and protons ($N = Z = 20$). ^{40}Ca is the heaviest stable nucleus to have an equal number of neutrons and protons. This stability can be attributed to its doubly magic nature. To remain stable, nuclei heavier than ^{40}Ca must have an excess of neutrons to compensate for the increasingly disruptive long-range electrostatic repulsion between the protons.

Doubly magic nuclei with $N = Z$ should be the ideal testing ground for the nuclear shell model because, to a good approximation, the neutrons and protons occupy the same quantum states. As their wavefunctions overlap, they tend to reinforce one another's structures, and hence shell effects should be amplified. Unfortunately, only two such nuclei exist beyond ^{40}Ca ; they are ^{56}Ni ($N = Z = 28$) and ^{100}Sn ($N = Z = 50$). ^{100}Sn lies very far from stability, close to the proton drip-line—the point at which nuclei can spontaneously emit a proton from their ground states². It was discovered three years ago at the GSI and GANIL laboratories, by fragmenting heavier nuclei^{3,4} fired at fixed targets. These experiments demonstrated the existence of the ^{100}Sn nucleus, but did not yield insight into its structure.

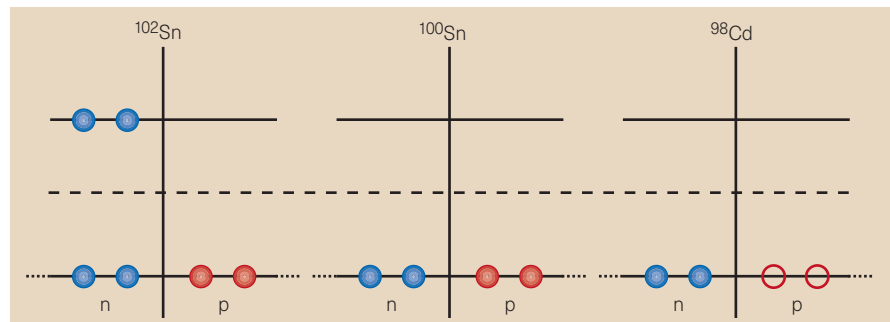


Figure 1 The neutron and proton shell occupancies of ^{102}Sn and ^{98}Cd in relation to a core of ^{100}Sn , an unstable but symmetrical nucleus with 50 protons and 50 neutrons. ^{102}Sn has two extra neutrons, which must sit at higher energies in the next shell; ^{98}Cd has two proton 'holes' (hollow circles). The opposite behaviours of ^{98}Cd and ^{102}Sn hint at an unexpected breakdown of neutron–proton symmetry.

For such insight, one ideally needs to study gamma-ray cascades from excited states of the nucleus. This is not yet feasible for ^{100}Sn , but new experiments^{1,5} at the Niels Bohr Institute in Risø have identified gamma-ray transitions in the neighbouring nuclei ^{98}Cd and ^{102}Sn . These nuclei are produced by firing ^{58}Ni ions at ^{46}Ti and ^{50}Cr targets, respectively. The nuclei fuse, and the compound nucleus then emits light particles, producing a range of product nuclei in excited states, each of which finally decays to its ground state by emitting gamma-rays. The experimenters exploit the relatively long lifetime of one of the gamma-decaying states of ^{98}Cd and ^{102}Sn by positioning their gamma-ray detection system around a catcher foil downstream of the target. This eliminates background gamma-rays emitted from more abundantly produced isotopes at the target position, as most of these isotopes will have gamma-decayed to their ground states before they reach the foil.

The gamma-ray transition energies of ^{102}Sn can be calculated by considering it as a pair of interacting valence neutrons coupled to a ^{100}Sn core. Similarly, excited states of ^{98}Cd can be viewed theoretically as two proton holes (like the electron holes of semiconductors) coupled to a ^{100}Sn core. But the observed gamma-ray energies from ^{102}Sn and ^{98}Cd (and other neighbouring nuclei with N or $Z = 50$) are similar to those from nuclei around the doubly magic nucleus ^{56}Ni (ref. 6), which suggests that the first excited state in ^{100}Sn is lower than theoretical expectations.

The lifetime of the long-lived gamma-decaying state in ^{98}Cd measured by Górska *et al.*¹ is also significantly longer than that expected from studies of neighbouring isotopes with $N = 50$. The lifetime of such states can be calculated in terms of an effective electric charge carried by the proton holes. This notion takes into account the nuclear-force interaction between the core and valence particles/holes, which has the effect of polarizing the core. Recent theoretical calculations, and the behaviour of similar states in neighbouring nuclei, imply that the holes in ^{98}Cd should have an effective charge much greater than the proton charge, and therefore increase the gamma-ray decay rate. But what is observed indicates an effective charge less than that of the proton. This very surprising result is compounded by the fact that the corresponding gamma-decay rate in ^{102}Sn is faster than expected, implying an unexpectedly high neutron effective charge in the neutron- ^{100}Sn core interaction.

This divergence in behaviour of proton-hole and neutron states may indicate a breakdown in the symmetry of neutron and proton shell structures near ^{100}Sn . This could be caused by the large Coulomb force or the loosely bound nature of the protons. The conundrum may be solved by further detailed

studies of core excitations in the vicinity of ^{100}Sn , and by the incorporation of additional nuclear structure effects such as the influence of the proton drip-line into theoretical calculations. But it is clear that the nucleus has not lost its capacity to surprise us. □

Philip Woods is in the Department of Physics and

Signal transduction

Inositol lipid pathways turn turtle

Kath Hinchliffe and Robin Irvine

It seems that there are few limits to the number of molecules and biochemical pathways into which evolution has incorporated *myo*-inositol. The awesome rate of new inositol phosphates identified in the 1980s was succeeded in the 1990s by a more sedate growth in the number of inositol lipids identified following the discovery of 3-phosphorylated inositol lipids (reviewed in ref. 1). Most people would have thought that the roster of these inositol lipids, and their metabolic interrelationships, is now complete. But two papers^{2,3} elsewhere in this issue tell us that we don't have all the answers yet.

All inositol lipids are derived from phosphatidylinositol (PtdIns). Subsequent phosphorylation of the hydroxyl groups at positions 3, 4 or 5 of the PtdIns inositol head group results in a large number of different phospholipids (Fig. 1), many of which are components of the lipid bilayers of cell membranes. By common consent, phosphatidylinositol-4,5-bisphosphate (PtdIns-4,5- P_2) is the key player because it is the precursor of at least three second-messenger molecules¹: inositol-1,4,5-trisphosphate (which modifies intracellular calcium levels), diacylglycerol (which regulates some members of the protein kinase C family) and phosphatidylinositol-3,4,5-trisphosphate (which is involved in signal transduction; see page 116 of this issue). In addition to this precursor role, PtdIns-4,5- P_2 may also be involved in several other cellular processes, including exocytosis, cytoskeletal regulation and intracellular trafficking of vesicles.

PtdIns-4,5- P_2 was thought to be produced exclusively from phosphatidylinositol-4-phosphate (PtdIns-4-P), and radiolabelling studies on whole cells and various membrane preparations have confirmed that this is probably the principal route of synthesis (Fig. 2). However, Rameh and co-workers² (page 192 of this issue) now demonstrate that there is an alternative pathway of synthesis for this molecule, which involves phosphorylation of the hydroxyl group at position *D*-4 of phosphatidylinositol-5-phosphate (PtdIns-5-P), a previously unrecognized inositol lipid (Fig. 2). What will amaze those in the field is that this step is catalysed by type II PtdIns-4-P 5-kinase, an enzyme that was purified in 1989 (ref. 4) on the basis of its apparent ability to produce

Astronomy, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK.

1. Górska, M. *et al.* *Phys. Rev. Lett.* **79**, 2415–2418 (1997).
2. Woods, P. J. *Nature* **381**, 25 (1997).
3. Schneider, R. *et al.* *Z. Phys. A* **348**, 241–242 (1994).
4. Lewitowicz, M. *et al.* *Phys. Lett. B* **332**, 20–24 (1994).
5. Lipoglavsek, M. *et al.* *Z. Phys. A* **356**, 239–240 (1996).
6. Grawe, H. *et al.* *Prog. Part. Nucl. Sci.* **38**, 15–28 (1997).

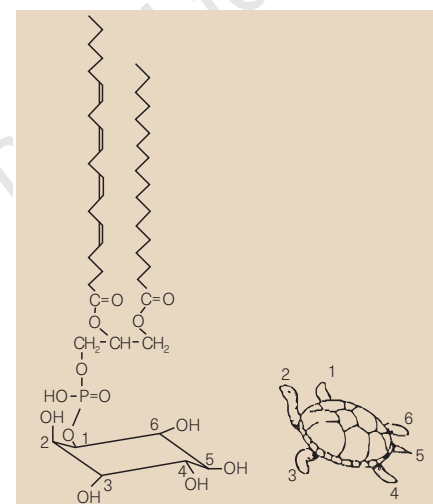


Figure 1 Structure of phosphatidylinositol (PtdIns). The structure depicts the *myo*-inositol ring (bottom) attached via a diester phosphate to a diacylglycerol moiety (top). As an *aide-mémoire* to the numbering of the carbons in the *myo*-inositol ring, we show 'Bernie Agranoff's Turtle'¹¹. The turtle's head is the 2-hydroxyl group, and the turtle is holding on to the diester phosphate (which links it to the diacylglycerol moiety) with its front right flipper (1-hydroxyl in *D*-numbering). The various phosphorylations of its two left flippers and tail are shown in Fig. 2.

PtdIns-4,5- P_2 from PtdIns-4-P. It now seems that this particular kinase is only active in these assays because of PtdIns-5-P contamination of PtdIns-4-P preparations; PtdIns-4-P is, in fact, a poor substrate for the enzyme, suggesting that it should be re-christened PtdIns-5-P 4-kinase (PIP4K). In support of other findings reported earlier this year⁵, Rameh *et al.* show that PIP4K can also phosphorylate PtdIns-3-P (an inositol lipid important for vesicle trafficking) to PtdIns-3,4- P_2 .

This then begs the question of whether the *in vivo* substrate of PIP4K is predominantly PtdIns-3-P or PtdIns-5-P. Little is known about cellular PtdIns-5-P levels and how they change when cells are stimulated with, for example, hormones and neurotransmitters, although Rameh and colleagues² report that resting mouse fibroblasts contain comparable levels of PtdIns-3-P and PtdIns-5-P. They also demonstrate that a small amount of PtdIns-5-P can be generated from the action

NATURE

100 YEARS AGO

Dr. Brunton is never content with a mere induction from experience when it is possible to suggest a rational explanation.

"I was once demonstrating the action of ammonia before a class here many years ago, and showed that if you held either ammonia or chloroform before the nose of a rabbit the heart stopped instantaneously. ... After the lecture was over, a student came up to me and said: 'If ammonia held before the nose stops the heart, how is it that it is of use in fainting? It ought to be exceedingly bad in fainting, and yet everybody knows it is good.' Well, I simply did not know. I said: 'I think it may possibly be that it tends at the same time to cause a deep inspiration, and thus stimulates the heart indirectly.' But I was not satisfied with this explanation, and so I put the question to the test of experiment. I found the answer to be this: At the same time that you stop the heart through the vagus by ammonia or any other irritating volatile substance held before the nose, you stimulate reflexly the vasomotor centre, cause contraction of the arterioles, and raise the blood-pressure enormously."

From *Nature* 11 November 1897.

50 YEARS AGO

Founder of modern British epidemiology, Charles Creighton ... was the most learned British medical scholar of the nineteenth century, who stood for something fundamental in the intellectual world of his generation. He was handicapped, however, by a curious mental obliquity which forced him to believe that the generally accepted must be necessarily false. Incapable of sharing much of the current medical teaching and frequently clashing with his medical colleagues, he brought matters to a head when in his article on vaccination in the "Encyclopædia Britannica" he declared that cowpox had nothing to do with smallpox, affording no protection against it. This was acclaimed by the antivaccinators, and Creighton was ostracized by the medical profession. He spent his last years in philosophical isolation in a tumble-down cottage at Upper Boddington, turning to Shakespeare for solace.

From *Nature* 15 November 1947.

of the SH₂-containing inositol 5-phosphatase on PtdIns-4,5-P₂ *in vitro*. But if this is the sole route of PtdIns-5-P production (that is, if there is no PtdIns-5-kinase activity in cells — and we cannot be sure of that yet), then the primary function of PIP4K would seem to be the removal of PtdIns-5-P, rather than the synthesis of PtdIns-4,5-P₂.

It is known that the product of PIP4K-mediated PtdIns-3-P phosphorylation, PtdIns-3,4-P₂, occurs physiologically, but its possible functions are still the subject of controversy. Sometimes it is viewed merely as a breakdown product of PtdIns-3,4,5-P₃ — an inositol lipid essential for growth-factor signalling mediated by the serine threonine kinase Akt⁶ (see page 116 of this issue). However, PtdIns-3,4-P₂ does have a direct role in the activation of Akt *in vitro*^{6,7} and the possibility remains that it has functions distinct from those of PtdIns-3,4,5-P₃. Obviously, a route for PtdIns-3,4-P₂ synthesis that does not involve PtdIns-3,4,5-P₃ would have interesting implications for the regulation of the Akt signalling pathway, and for 3-phosphorylated inositol lipids in general. So it may be that the major physiological function of PIP4K will turn out to be the generation of PtdIns-3,4-P₂.

The previously described (formerly only) route of PtdIns-4,5-P₂ synthesis from PtdIns-4-P (Fig. 2) is catalysed by the type I PIP kinases, which share homology with PIP4K. Rameh *et al.* confirm that these enzymes are indeed 5-kinases, and that they can phosphorylate PtdIns-3-P *in vitro*, producing yet another inositol lipid, PtdIns-3,5-P₂ (probably misidentified as PtdIns-3,4-P₂ in similar studies³). This molecule has been previously shown to occur in a fibroblast cell line⁸, and the radiolabelling data in that study suggested a route of synthesis by phosphorylation of the D-5 hydroxyl group on PtdIns-3-P.

So, does PtdIns-3,5-P₂ have a physiological function? A companion paper in this issue by Dove *et al.*³ (page 187) suggests that it does. These investigators show that osmotic stress in yeast leads to increased production of PtdIns-3,5-P₂. They suggest that this inositol lipid may either constitute a specific signal or be involved in some process that is modified under conditions of osmotic stress. They also provide further evidence for a widespread distribution of PtdIns-3,5-P₂ among eukaryotes, confirming the earlier observation⁸ that it is present in mammalian cell lines, and they also demonstrate its presence in plants. Moreover, they show that PtdIns-3,5-P₂ levels in fibroblasts decrease in response to hyperosmotic shock treatment, which contrasts with the increased levels seen in osmotically stressed yeast.

The substrate for the (as yet unidentified) PtdIns-3-P 5-kinase is synthesized in yeast by a protein encoded by the *vps34* gene⁹. Mutations in this gene produce defects in vesicle-mediated protein sorting¹⁰, suggesting that PtdIns-3-P plays an important role

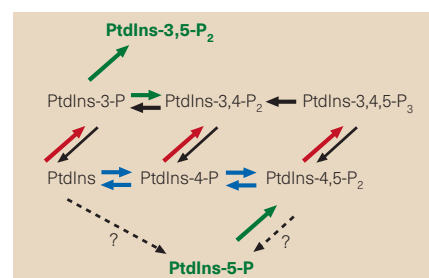


Figure 2 Synthesis of inositol phospholipids. All inositol phospholipids have phosphatidylinositol (PtdIns) as a precursor. Subsequent phosphorylation of the hydroxyl groups at positions 3, 4 or 5 of the inositol head group provides a variety of inositol phospholipids, all of which have now been found to occur naturally. Arrows from a more highly phosphorylated lipid towards a less highly phosphorylated derivative denote the action of a phosphatase. Blue arrows show the interconversion of the best-known inositol lipids: PtdIns, PtdIns-4-P and PtdIns-4,5-P₂. Red arrows denote the production of the 3-OH phosphorylated lipids (at least *in vitro*) catalysed by the PtdIns-3-kinase family of enzymes. Green indicates the newly described inositol lipids, PtdIns-5-P and PtdIns-3,5-P₂, and the synthesis pathway from PtdIns-3-P to PtdIns-3,4-P₂. The dashed arrows represent unknown pathways of PtdIns-5-P synthesis.

in this pathway. However, the existence of another inositol lipid, PtdIns-3,5-P₂, further downstream of Vps34p activity, means that the involvement of PtdIns-3-P in vesicle trafficking may be indirect, and that PtdIns-3,5-P₂ is actually the functionally important inositol lipid in this system.

So, the six cellular inositol lipids already recognized can be joined by another two: PtdIns-5-P and PtdIns-3,5-P₂. The search is now on to identify the functions of these newly described inositol phospholipids, the molecules with which they interact and — at least for PtdIns-5-P — their mechanisms of synthesis. In previous studies they have been overlooked, doubtless because their detection depends on careful analysis by high-performance liquid chromatography. So we can't help but wonder, are there any more out there? □

Kath Hinchliffe and Robin Irvine are in the Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, UK.

1. Divecha, N. & Irvine, R. F. *Cell* **80**, 269–278 (1995).
2. Rameh, L. E., Tolias, K. F., Duckworth, B. C. & Cantley, L. C. *Nature* **390**, 192–196 (1997).
3. Dove, S. K. *et al.* *Nature* **390**, 187–192 (1997).
4. Ling, L. E., Schulz, J. T. & Cantley, L. C. *J. Biol. Chem.* **264**, 5080–5088 (1989).
5. Zhang, X. *et al.* *J. Biol. Chem.* **272**, 17756–17761 (1997).
6. Stokoe, D. *et al.* *Science* **277**, 567–570 (1997).
7. Franke, T. F. *et al.* *Science* **275**, 665–668 (1997).
8. Whiteford, C. A., Brearley, C. A. & Ulug, E. T. *Biochem. J.* **284**, 39–45 (1997).
9. Schu, P. V. *et al.* *Science* **260**, 88–91 (1993).
10. Herman, P. K. & Emr, S. D. *Mol. Cell Biol.* **10**, 6742–6754 (1990).
11. Agranoff, B. W. *Life Sci.* **32**, 2047–2054 (1983).

Photonics

The Vs and Qs of optical microcavities

Pauline Rigby and Thomas F. Krauss

Periodic structures can be used to manipulate light. Such structures are called photonic bandgap materials, or photonic crystals¹, and promise to be as important to the progress of optical devices as semiconductors were to the development of electronic devices. Physicists didn't have a good 'box' for optical radiation (light with wavelengths in or near the visible region) before, because metals are not perfect reflectors at optical wavelengths. But photonic crystals can achieve very strong confinement of optical light. On page 143 of this issue², Foresi *et al.* of MIT report the most spatially localized photon mode ever, which was achieved using the photonic bandgap approach. The volume of their 'box' is $0.055 \mu\text{m}^3$, which is only five cubic half-wavelengths ($1.54\text{-}\mu\text{m}$ infrared light being shortened to $0.44 \mu\text{m}$ by the material's refractive index of 3.5). When the box is this small, the radiation properties of the material inside it are fundamentally altered³.

The work adds to the limited experimental evidence that photonic crystals do what they are supposed to do. The properties of these materials are thought to arise from multiple reflections at the periodically distributed interfaces, which add up to prevent light from propagating over a wide band of wavelengths. Although the transmission properties of photonic crystals have been measured before⁴, the mechanism had not been demonstrated so conclusively.

Introducing a cavity into the structure has made it possible to prove that reflection, rather than diffraction or other loss mechanisms, is responsible for the opaque spectral regions — an important distinction because it means that a cavity must have reflective walls to trap light. This method has also been used in larger cavities with two-dimensional photonic crystal boundaries⁵, which clearly demonstrate the simultaneous occurrence of low transmission and high reflection⁶.

The structure investigated by the MIT group⁷ is periodic in one dimension, consisting of a line of cylindrical holes etched through a ridge of silicon (Fig. 1a). The ridge is a waveguide, which confines light in two dimensions by total internal reflection at dielectric boundaries. In the third dimension (along the ridge) light is contained by the periodic structure. The first waveguide-based microcavity to have its spectral properties fully characterized⁸ was also a one-dimensional grating (Fig. 1b), which showed strong confinement but leaked into the substrate. By instead using a design where the

semiconductor material is placed on a glass substrate, the MIT researchers have created a structure with very low optical losses.

Foresi *et al.* created their cavity by introducing a single larger gap between holes in the middle of the line. This produces a very sharp transmission peak within the region of low transmission, where the cavity resonance occurs. So the device acts as an extremely narrow, wavelength-selective filter.

The ability of the cavity to confine radiation is described by the quality factor or 'Q', a measure of the sharpness of the transmission peak (the ratio of the width of the peak to its frequency). It is also the ratio of the optical power stored in the cavity to the cycle-average power radiated out of the cavity. Therefore the larger Q is, the longer radiation is trapped. Q and the volume V can be combined to create a figure of merit for microcavities, called the Purcell-number³ or the spontaneous emission enhancement factor. A high value of Q/V is desirable.

The MIT team have achieved this by making the volume tiny — $0.055 \mu\text{m}^3$ — which should give them a spontaneous emission enhancement factor of 34; that is, a radiating atom inside the cavity should generate light 34 times faster than it would

without the cavity, because of the resonant coupling. Brighter, more power-efficient devices could result. But, more importantly, devices such as laser diodes or light-emitting diodes could be modulated at a higher frequency, and so transmit more bits per second, increasing the capacity of optical communication systems.

This cannot be tested using the new structure, because it doesn't emit light, but the idea has been demonstrated with a pillar-type microcavity⁹ (Fig. 1c), for which the trapped mode of light had a Purcell-number of 4. Instead of having a tiny volume, this device uses quantum dots to provide a high-Q light source (J.-M. Gerard, personal communication). When the quantum dots were excited, they emitted light four times faster than they would have done without the cavity, as predicted.

To take full advantage of the high Purcell-number of their cavity, the MIT team will have to find a material with a Q that is comparable to the cavity Q. The problem is that although the cavity can increase the material's emission if the emission line and the resonance overlap, it can also suppress the emission if they don't. So if only a small part of the material's emission line overlaps with the cavity resonance (as happens when the Qs are very different), not much is gained. Finding such a high-Q material is not straightforward, considering that the material Q of standard light-emitting semiconductors is between 20 and 30, whereas the measured Q of the cavity is 265.

If the new structure is to develop into a commercial device for high-speed communications, Foresi *et al.* will also have to consider the other properties of their cavity, such as its luminescence and electrical conductivity. And they must find a way to inject current, as the glass substrate now used will not conduct. But it may be possible to combine an all-semiconductor approach with the MIT design to create a new source of light for ultra-fast communication systems. □

Pauline Rigby is in the Department of Engineering Science, Oxford University, Parks Road, Oxford OX1 3PJ, UK.

e-mail: pauline.rigby@eng.ox.ac.uk

Thomas F. Krauss is in the Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8QQ, UK, currently on leave at California Institute of Technology, Mail stop 136-93, Pasadena, California 91125, USA.

e-mail: tfk@cco.caltech.edu

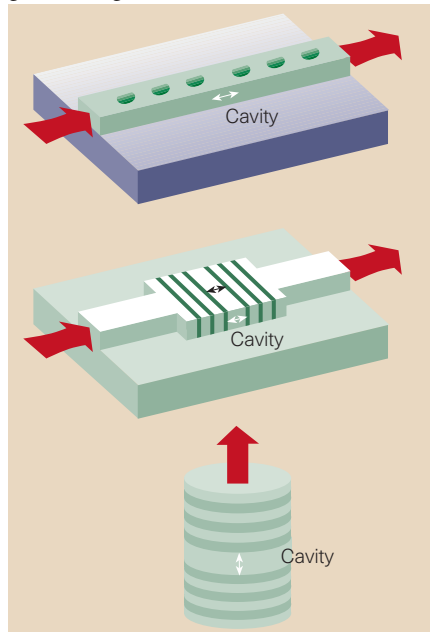


Figure 1 Three traps for photons: a, the smallest optical microcavity yet built, formed by a larger gap in a periodic series of holes in a silicon waveguide²; b, an earlier, grating-like microcavity⁸; c, a cylindrical cavity, used to demonstrate the acceleration of optical emission⁹. (Not to scale.)

- Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. *Nature* **386**, 143–149 (1997).
- Foresi, J. S. *et al.* *Nature* **390**, 143–145 (1997).
- Purcell, E. M. *Phys. Rev.* **69**, 681–681 (1946).
- Krauss, T. F., De La Rue, R. M. & Brand, S. *Nature* **383**, 699–702 (1996).
- Labilloy, D. *et al.* *Electron. Lett.* **33**, 1978–1979 (1997).
- Labilloy, D. *et al.* *Phys. Rev. Lett.* (in the press).
- Villeneuve, P. R. *et al.* in *Photonic Band Gap Materials* (ed. Soukoulis, C. M.) 411–426 (NATO ASI Series, Kluwer, Dordrecht, 1996).
- Krauss, T. F., Vögeler, B., Stanley, C. R. & De La Rue, R. M. *IEEE Photonics Technol. Lett.* **9**, 176–178 (1997).
- Gerard, J.-M. *et al.* *Appl. Phys. Lett.* **69**, 449–451 (1996).

An interview with Daedalus

This is the 500th contribution from Daedalus to appear in News and Views. To mark the occasion, the item is not his usual column, but an interview with Daedalus himself.

NATURE: The first account of your research activities published in *Nature* was back in 1988. Your arrival caused us some consternation, as I recall.

DAEDALUS: It was a bold move for both of us! But of course my career goes back much further. It began in 1964, and until 1988 was reported weekly in *New Scientist*.

NATURE: As we know from the letters we receive, many of our readers are impressed by the wide range of your research activities. Do any particular highlights stand out?

DAEDALUS: My finest hour was undoubtedly in 1966, when I proposed the class of carbon molecules later called fullerenes¹. At that time there was no good way of making them. In 1982 I outlined my guesses at their character and geodesic properties². In 1986, of course, the first of the class, buckminsterfullerene, was synthesized and identified. Last year it gained for its discoverers the Nobel Prize for Chemistry. They and many other workers in this rapidly expanding field have generously acknowledged my contributions³⁻⁵. I still maintain an interest in the topic⁶. I was the first to suggest what were later called carbon nanotubes⁷ (though of a larger diameter than those currently being studied), and am among those who proposed the graphite foam structure^{8,9}.

A second early prediction of mine, in 1968, was the chemical laser¹⁰. I didn't know that Kasper and Pimentel were working on the same lines and were ahead of me. Even so, their system at that time required an initial photodissociation¹¹. My admittedly naive proposal for a nitroglycerine laser envisaged the ultimate form, a purely chemical system, and that step was only taken in 1969 by Cool and Stephens¹². And certainly in that scheme I was the first to suggest the use of a chemical laser to shoot down ballistic missiles. The idea got only a small mention in the press at the time; but Ronald Reagan, as I recall, later took it up in a big way.

Another highlight was a deduction from the well-known claim that the right side of the brain is emotional, whereas the left side is rational. In 1991 I suggested that, by repeatedly breathing in through one nostril, one could cool the hemisphere above that nostril into quiescence. The other hemisphere, warmed by repeated breathing out, would become dominant. Thus the breather could bias his psychology to the task in hand¹³. It

turned out that I had anticipated an independent paper¹⁴ in the same field which in 1995 was awarded the coveted Ig-Nobel prize. So I am well represented among the award-winning strands of modern research.

NATURE: Has your work had an impact in any other fields?

DAEDALUS: I like to think so, but literature references are sparse. It is possible that some scientists do not take my contributions entirely seriously. However, my proposal¹⁵ of 1990 to synthesize benzene in a magnetic field, so as to thread the molecules permanently around the lines of force, has since been given a very interesting theoretical treatment¹⁶, and discussed in the fullerene context¹⁷. My deduction of the existence of toroidal bubbles¹⁸ has been verified by several workers^{19,20}. It turns out that they had already been observed in the exhalations of whales and dolphins²¹. My scheme to invert the principle of X-ray tomography to optimize radiation therapy and plastic moulding²² has been cited by workers active in the field^{23,24}.

NATURE: So you keep a careful watch on the literature?

DAEDALUS: Not at all. I ignore it almost entirely; it just cramps one's originality. If you have an idea for an experiment or a calculation, try it, I say, and worry about the literature afterwards! Sometimes, of course, this policy backfires, and I find that I've been anticipated. For example, in 1996 I suggested²⁵ a novel use for heavy water — to counter mad-cow disease. I reckoned it could refold and reform the rogue prions responsible. A reader promptly told me that his team had already tried it. It doesn't seem to work with prions, but it can correct the folding defects of some other proteins²⁶. Again, early this year I proposed²⁷ the existence of 'Carnot creatures' living around superheated oceanic 'black smoker' springs, and gaining their energy thermodynamically as heat-engines, by exploiting the ambient temperature gradients. I even wondered if life itself might have evolved originally as a form of heat-engine, at least on other planets. A reader rapidly made me aware of a speculation that early organisms were indeed Carnot creatures²⁸. Later it turned out that many others have also proposed the origin of life in black smokers²⁹.

NATURE: You mention readers' responses. Do our readers play much part in your research?

DAEDALUS: Yes indeed. One great advantage of publicity in *News and Views* is the feedback I receive. This is particularly true when a column is reported more widely in other media. When I proposed a way to determine the mass, spin and moment of inertia of the soul³⁰, the scheme received wide publicity. I had a vast postbag of letters, some proposing values for these quantities. Again, it was a reader who informed me of a failed nineteenth-century Dutch attempt to grow coca bushes in the East Indies³¹, which led to my proposal to solve the cocaine problem by spreading an infectious plant virus to inhibit the coca bush from making the drug³². This was reported in some detail, and attracted the interest of a well-known biotechnology company. I like to think that somewhere a drug enforcement agency is now trying it in practice.

Your readers also provide more humbling forms of feedback. They rap me sharply over the knuckles when I get things wrong — as in my ideas about propagating light inside metals^{33,34}, rigging the ovulation contest between ovaries as a method of contraception^{35,36}, and comets accumulating solid hydrogen in the outer regions of their eccentric orbits^{37,38}.

Even when I'm right, they sometimes tell me that I'm well behind the times. When I outlined a method for making metal foam³⁹, not only did I receive a letter from an organization that was already doing it; another reader sent me two actual specimens of metal foam. My scheme for an anti-theft device based on a short-range radio transmitter⁴⁰ brought an indignant complaint from a company already making one. A reader later sent me an actual sample of a consumer device based on this principle. Worse, it was a remaindered product, being sold off cheaply. Just the other day I had a letter from a reader whose research group, quite independently, was studying the composition and calming effect on children of amniotic fluid⁴¹ — a project I outlined in September⁴².

Indeed, through your columns DREADCO has had requests for samples or specifications of products and requests to manufacture them under licence, job applications, invitations for consultancy, offers to take part in research trials — and even threats of legal action for patent infringement.

NATURE: Ah yes, DREADCO. Tell us about this research organization of yours.

DAEDALUS: Certainly. DREADCO, the Daedalus Research Evaluation And Development Corporation, to give it its full title, is the R&D company through which I develop my more complicated ideas. In my role as research director, I maintain a firmly contrarian policy. Most industrial research these days is throttled by report-bound management

theorists and unimaginative bean-counters: mismanagement by objectives, as the saying goes. Our contrarian stance is therefore one of extreme speculation and irresponsibility — a very neglected niche in the modern R&D landscape. Our staff are the sort of creative troublemakers who cannot tolerate the regimented, conformist, career-driven ambience of more respectable laboratories, and I give them the utmost freedom. Most of our projects fail, of course — but this is true even in the most obsessively regulated R&D outfits. But we rapidly learn from our failures, and hasten to build on the unexpected findings which our undisciplined researchers keep stumbling over. When we find a promising product on our hands, we may go in for a little pilot manufacturing and test marketing, mainly for evaluation purposes, before selling it off for exploitation to some other, more buttoned-down company.

NATURE: And how do you choose your research topics?

DAEDALUS: Again, by taking good care to ignore what other people are doing. For example, your pages are dominated by protein-botherers and gene-tweakers. Your product advertisers think of almost nothing else. It's a massively crowded and competitive area. And what practical advances or new products have come out of it? Relative to the effort going in, almost none. The other currently fashionable technical mania is

information technology. I almost weep to think of all the potentially fruitful, creative minds who get hooked on compulsive computer nerdery, and are thenceforth lost to civilization. Analysts look in vain for any increase in productivity or economic value from all this digital churning. So, following our contrarian philosophy, at DREADCO we do very little of either. Of the projects you have publicized, only about 2% concern these over-subscribed fields.

NATURE: Do you have any message or advice for our readers?

DAEDALUS: I hesitate to lecture such an audience, especially in view of my own patchy record of success — only a small proportion of my ideas ever get much further than your columns. But even Sir Peter Medawar, who won the 1960 Nobel Prize for Medicine or Physiology for his work on immunology, once remarked that for all the good it had been to science, about 80% of his research activity had been completely wasted. He felt that this was quite a typical record; many of your readers may look back ruefully on a similar percentage of fruitless research effort. An activity which is only about 20% efficient should not be taken too seriously. So it's worth exploring the odd crazy notion or trying the occasional mad experiment, if only to get a bit of fun as a by-product. Indeed, it might seem optimal to use 80% of your time this way. I recommend it.

David Jones

1. *New Scientist* 245, 3 Nov. (1966).
2. Jones, D. *The Inventions of Daedalus* 118 (Freeman, Oxford, 1982).
3. Smalley, R. E. *Angew. Chem. Int. Edn Engl.* **36**, 1594 (1997).
4. Bakowies, D., Bühl, M. & Thiel, W. *J. Am. Chem. Soc.* **11**, 10113 (1995).
5. Ballone, P. & Milani, P. *Phys. Rev. B* **42**, 3201 (1990).
6. Jones, D. *Phil. Trans. R. Soc. Lond. A* **343**, 9 (1993).
7. *New Scientist* 1, 24 Apr. (1986); 80, 1 May (1986).
8. Jones, D. *Nature* **351**, 526 (1991).
9. MacKay, A. L. & Terrones, H. *Nature* **352**, 762 (1991).
10. *New Scientist* 615, 19 Sep. (1968).
11. Kasper, J. V. V. & Pimentel, G. C. *Phys. Rev. Lett.* **14**, 352 (1965).
12. Cool, T. A. & Stephens, R. R. *J. Phys. Chem.* **51**, 5175 (1969).
13. Jones, D. *Nature* **349**, 198 (1991).
14. Shannahoff-Khalsa, D. S., Boyle, M. R. & Buebel, M. E. *Int. J. Neurosci.* **57**, 238 (1991).
15. Jones, D. *Nature* **348**, 588, 682 (1990).
16. Kusmartsev, F. V. *Phys. Lett. A* **161**, 433 (1992).
17. Haddon, R. C. *Nature* **388**, 31 (1997).
18. Jones, D. *Nature* **343**, 122 (1990).
19. Spisak, W. *Nature* **349**, 23 (1991).
20. Semper, R. J. *Nature* **356**, 390 (1992).
21. Aref, H. *Nature* **356**, 200 (1992).
22. Jones, D. *Nature* **343**, 516 (1990).
23. Webb, S. *Nature* **344**, 284 (1990).
24. Webb, S. *Phys. Med. Biol.* **36**, 1201 (1991).
25. Jones, D. *Nature* **381**, 118 (1996).
26. Randell Brown, C. et al. *Cell Stress & Chaperones* **1**, 117 (1996).
27. Jones, D. *Nature* **285**, 301 (1997).
28. Muller, A. W. J. *Proc. Biophys. Mol. Biol.* **63**, 193 (1995).
29. Russell, M. J. & Hall, A. J. *J. Geol. Soc.* **154**, 377 (1997).
30. Jones, D. *Nature* **366**, 412, 516 (1993).
31. Seitz, R. (personal communication).
32. Jones, D. *Nature* **339**, 514 (1989).
33. Jones, D. *Nature* **379**, 124 (1996).
34. Kittel, C. *Nature* **380**, 282 (1996).
35. Jones, D. *Nature* **362**, 502, 594 (1993).
36. Jansen, R. *Nature* **367**, 10 (1994).
37. Jones, D. *Nature* **356**, 482 (1992).
38. Weissman P. R. *Nature* **358**, 240 (1992).
39. Jones, D. *Nature* **373**, 564 (1995).
40. *New Scientist* 72, 16 Aug. (1984).
41. Marlier, L. Schaal, B. & Soussignan, R. *Child Dev.* (in the press).
42. Jones, D. *Nature* **389**, 238 (1997).

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

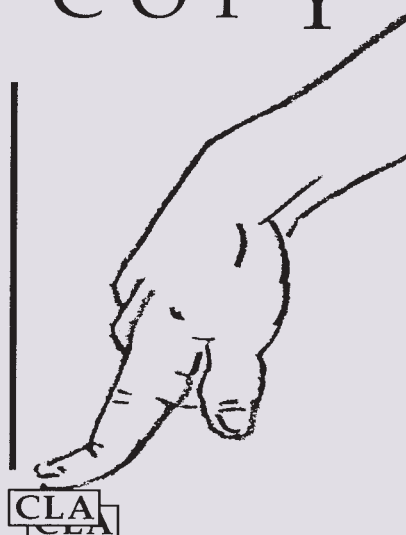
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

Piero's perspective

Renaissance painter Piero della Francesca plotted perspective so meticulously and consistently that he was able to make the divine, with its miraculous lack of optical logic, shine out in contrast to the rest of the picture.

Martin Kemp

Mention the interplay of art and science — or, more specifically, any instance of the influence of art on science — and perspective is likely to feature prominently. Discovered early in the fifteenth century by the great Florentine architect Filippo Brunelleschi, and first codified in writing by Leon Battista Alberti in his little book *On Painting* in 1435, perspective not only became the stock mode of representing any three-dimensional form in science and technology, but also played a key role in the development of projective geometry around 1600 and of descriptive geometry two centuries later.

For many artists, perspective assumed the status of a convenient technique — a routine geometrical-cum-optical trick to establish the illusion of a space behind the surface of a picture. But for a minority of theoretically inclined artists it acted as a research tool, through which the picture became an experimental field for the geometrical construction of illusions emulating the optical force of real space.

Even more exceptionally, the mastery and use of optical rules could constitute an ethical imperative. The greatest of such artists was the Renaissance painter, Piero della Francesca. Even a quick inspection of his relatively small painting of the *Flagellation* will reveal the meticulous construction of an exceptionally deep platform, passing back from the zone in which three figures debate the nature of Christ's torment, as disclosed in the middle ground, to a distant piazza. Careful analysis reveals the extremes to which he went to comply with mathematical rigour. For example, the complex tile pattern within the loggia is projected with scrupulous preci-



Piero della Francesca, *Flagellation*, c. 1465.

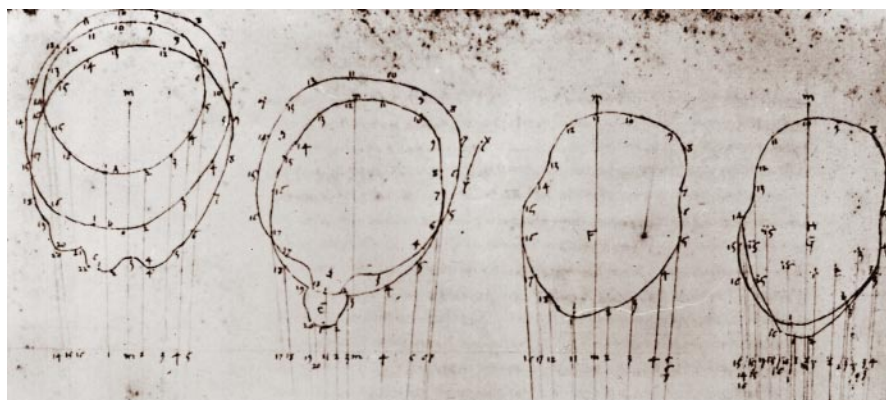
sion, and is founded upon a series of related squares, the larger of which have side lengths equal to the diagonal of the smaller (i.e. $\sqrt{2}$).

Piero wrote books on practical mathematics with simple algebra, on the five regular polyhedra, and, unsurprisingly, on *The Perspective of Painting*. His treatise on perspective shows how to plot the outlines of forms systematically on to a foreshortened plane, and how to use the plan and elevation of solid bodies to project them point-by-point on to an intersecting plane from a fixed viewpoint. The illustrations below show the first and final stages in the plotting of points disposed in horizontal bands around a head. The final outline results from the coordination of points projected from sections and profile. We know that he followed such meticulous procedures in wall paintings,

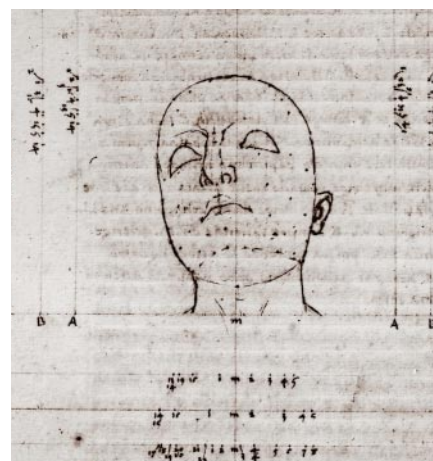
even when the effects would be so far away from the spectator that they could not be appreciated. Getting things right was integral to his contract with God's design of nature.

But, amid this extreme rationality, he was a painter of miracles, which lie outside optical logic. In the *Flagellation* the space occupied by Christ, and the section of ceiling above his column, are paradoxically flooded by a burst of light from a source within the architectural space. Only Christ is aware of this miraculous intrusion. It is the very logic of Piero's perspectival consistency that allows him to highlight the supra-normal power of the divine. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



Projection of successive horizontal sections of a head on to picture plane, from Piero della Francesca's *The Perspective of Painting*.



Submarine columns of ikaite tufa

In the small Ikka Fjord in southwestern Greenland, we have studied remarkable submarine tufa columns forming over alkaline springs by abiotic precipitation of the metastable, cold-water mineral ikaite ($\text{CaCO}_3 \cdot 6\text{H}_2\text{O}$)¹⁻³. The columns, which support a rich marine life, are up to 20 m high, with annual growth of more than 50 cm. We wish to nominate the locality, which seems to be unique, for classification as a geological World Heritage Site.

Ikka Fjord is a glacial valley, flanked by steep, 500-m-high, flat-topped mountains of Precambrian gneiss⁴. Syenitic and carbonatitic rocks of the 1,300-million-year-old Grønndal-Ika igneous complex cut the fjord in a 3-km-wide belt with a north-westerly trend⁵. The fjord water is marine except for the uppermost 1–2 m, where freshwater runoff lowers the salinity to <20‰. The ikaite columns are in the shallow, innermost part of the fjord (Fig. 1a). They are visible from the surface during periods of calm weather and at low tide as greenish-white, pointed or mushroom-shaped towers.

We observed the columns in the fjord by sub-aqua diving, side-scan sonar and acoustic profiling. Columns are restricted to a 0.75 km² 'garden' area, where there are more than 500 individual columns of 1–20 m in height. They form upright, trunk-like structures with diameters from a few centimetres to several metres (Fig. 1b). The area corresponds to shoreline outcrops of the Grønndal-Ika complex. Columns are rooted in bottom sediments, at the sides of rocky exposures or in massive, dome-shaped tufa build-ups.

Often the columns occur as clusters growing from a common base. Branching is rare except at the top of the highest columns, where metre-wide, saucer-shaped crowns may form close to the surface. Growth zones consist of white, fine-crystalline ikaite without algal overgrowth, which contrasts with the greyish-green, algal-covered older parts of the columns. Growth is generally confined to the top of the columns, but is also evident as finger-shaped extrusions elsewhere on the columns. Cross-sections and radiograms show that the inner structure is porous and sometimes contains an irregular conduit. Coralline red algae (*Lithothamnion* and *Clathromorphum*) encrust the lower part of the columns and thereby stabilize the delicate column structures.

Pauly⁶ has suggested that the columns form by seepage of fresh water from the bottom of the fjord. To test whether seepage takes place we cut eight columns one metre from their bases and fastened watertight

sample bags to the stumps. The bags were filled within days by water that was less dense than the surrounding sea water. In most cases, a precipitate of millimetre-sized ikaite crystals could be seen on the stumps. One stump, which we revisited after one year, showed vertical growth of more than 50 cm (Fig. 1b).

We took pore-water samples directly from columns with underwater syringes. The seep water is a sodium bicarbonate and sodium carbonate brine with high pH, high alkalinity and high phosphate content. The calcium concentration is significantly lower than in sea water (Table 1). The hydrogen and oxygen isotope composition of the water is close to that of lake and stream water sampled on the top of the intrusive complex. Cold homothermic springs (3.0–3.5 °C) surrounded by luxuriant plant communities, including three species of orchids, are found on the fjordward side of the intrusion. The springs have raised alkalinities (pH 8–9), in contrast to the nearly distilled waters of springs and brooks outside the intrusion area, and contain remarkable marine fauna elements such as salt mites (*Halacarida*).

The juxtaposition of ikaite columns, springs and carbonatites, and the total lack of columns outside the intrusive area, indicate a link. The highly fractured intrusive complex probably acts as an aquifer for groundwater originating from precipitation on top of the intrusion. Dissolution of sodium-rich carbonate minerals in the intrusion would increase the alkalinity and sodium content of the water. If such alkaline groundwaters mix with sea water that is high in calcium at seeps on the bottom of the fjord,

immediate supersaturation would lead to rapid precipitation of calcium carbonate.

The high phosphate content of seep water and low temperatures at seepage sites favour precipitation of ikaite rather than calcite or aragonite⁷. The elongated shape of the monoclinic ikaite crystals causes the formation of the permeable framework of the columns and thereby ensures their vertical growth. In this model, seeps literally

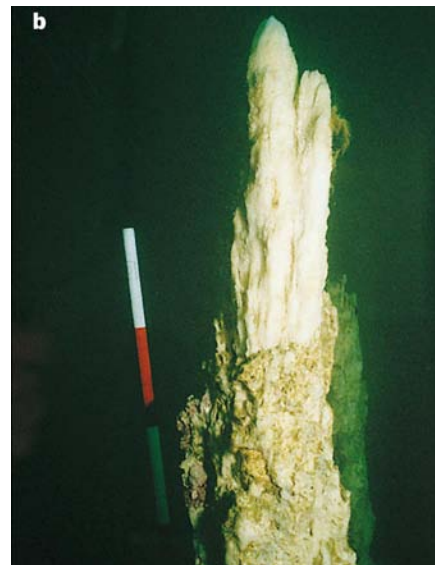
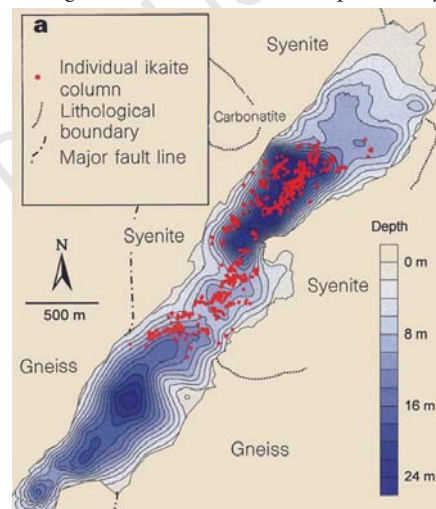


Figure 1 Distribution and typical shape of the ikaite tufa columns in Ikka Fjord. **a**, Bathymetric map of the inner part of the fjord, where locations of the largest columns are shown in red. Columns develop immediately beyond the shallow waters of the delta at the top of the fjord and occur continuously up to the regional fault line bisecting the fjord. Syenites and carbonatites belong to the Grønndal-Ika igneous complex. **b**, Ikaite column at 9 m water depth. Red and white subdivisions on the pole are 20 cm each. Newly formed ikaite above the oblique cutting surface represents 13 months of growth. Older column below the cut surface is covered with *Lithothamnion* and other algae.

Table 1 Composition of seep water compared with local sea water

Component	Seep water	Sea water
Conductivity (mS cm ⁻¹)	18.2	42.5
Salinity (‰)	9.3	31.1
Temperature (°C)	4.0	3.6
pH	10.4	8.1
Na ⁺ (mmol l ⁻¹)	198	413
K ⁺ (mmol l ⁻¹)	1.9	9.5
Ca ²⁺ (mmol l ⁻¹)	0.17	8.9
Mg ²⁺ (mmol l ⁻¹)	1.7	45.7
Cl ⁻ (mmol l ⁻¹)	21.2	506
SO ₄ ²⁻ (mmol l ⁻¹)	2.8	28.6
PO ₄ ³⁻ (mmol l ⁻¹)	0.26	below detection
Alkalinity (mmol l ⁻¹)	153	<0.5
δ ² H _{water} ‰ SMOW	-95.7	-10.4
δ ¹⁸ O _{water} ‰ SMOW	-13.4	-0.9

Seep water was obtained by underwater pumping of pore water from an active column at 10 m water depth. Seawater for comparison was sampled 1 m from the column. Conductivity, pH and alkalinity were measured in the field. Alkalinity is expressed as total alkalinity. Isotopic values are given as ‰ deviations from standard mean ocean water (SMOW).

create their own conduits in the form of vertical, chimney-like columns. Only the formation of sea-ice during the winter would prevent further upwards growth of the columns.

To our knowledge, the Ikka columns represent a phenomenon unique in the world. The now inactive tufa towers found at the shores of Mono Lake⁸ and Pyramid Lake^{9,10}, in the western United States, may represent structures formed in a similar way to the Ikka columns, but in non-marine environments. The high scientific and aesthetic value of the Ikka columns make them appropriate for international protection.

Bjørn Buchardt, Paul Seaman*
Gabrielle Stockmann, Marie Vous
Uffe Wilken

Geological Institute, University of Copenhagen,
Øster Voldgade 10, DK-1350 Copenhagen K,
Denmark *and Department of Geology,
Imperial College of Science, Technology and Medicine,
London SW7 2AZ, UK
e-mail: bjornb@geo.geol.ku.dk

Lene Düwel, Aase Kristiansen

Botanical Institute, University of Copenhagen,
Ø. Farimagsgade 2, DK-1353 Copenhagen K,
Denmark

Christopher Jenner

ETSU, Harwell Laboratories, Oxford OX11 0RA, UK

Michael J. Whitham

School of Earth and Ocean Sciences,
University of Victoria, PO Box 3050, Victoria,
British Columbia, V8W 2Y2 Canada

Reinhardt M. Kristensen

Godtfred H. Petersen, Lone Thorbjørn

Zoological Museum, Universitetsparken 15,
DK-2100 Copenhagen K, Denmark

1. Pauly, H. *Arctic* **16**, 263–264 (1963).
2. Marland, G. *Geochim. Cosmochim. Acta* **39**, 83–91 (1975).
3. Suess, E. *et al. Science* **216**, 1128–1131 (1982).
4. Berthelsen, A. & Henriksen, N. *Ivigut 61 V.1 Syd. Geological Map of Greenland 1:100,000* (Geol. Surv. Greenland, Copenhagen, 1975).
5. Emeleus, C.H. *Medd. Grønland* **172**(3), 1–75 (1964).
6. Pauly, H. *Naturens Verden* Argang 1964, 168–192 (Copenhagen, 1963).
7. Johnston, J., Merwin, H. E. & Williamson, D. E. *Am. J. Sci.* **41**, 473–493 (1916).
8. Bischoff, J. I. *et al. Geochim. Cosmochim. Acta* **57**, 3855–3865 (1993).
9. Shearman, D. J., McGugan, A., Stein, C. & Smith, A. J. *Geol. Soc. Am. Bull.* **101**, 913–917 (1989).
10. Benson, L. *Palaeogeogr. Palaeoclimatol., Palaeoecol.* **109**, 55–87 (1994).

Saccades without eye movements

When reading text, human subjects use a pattern of eye movements consisting of fast saccadic movements and fixations¹. We have found a subject who cannot make eye movements. Her visual perception is surprisingly normal and she is able to read at high speeds. She uses movements of the head to compensate for the absence of eye movements. Her head movements during

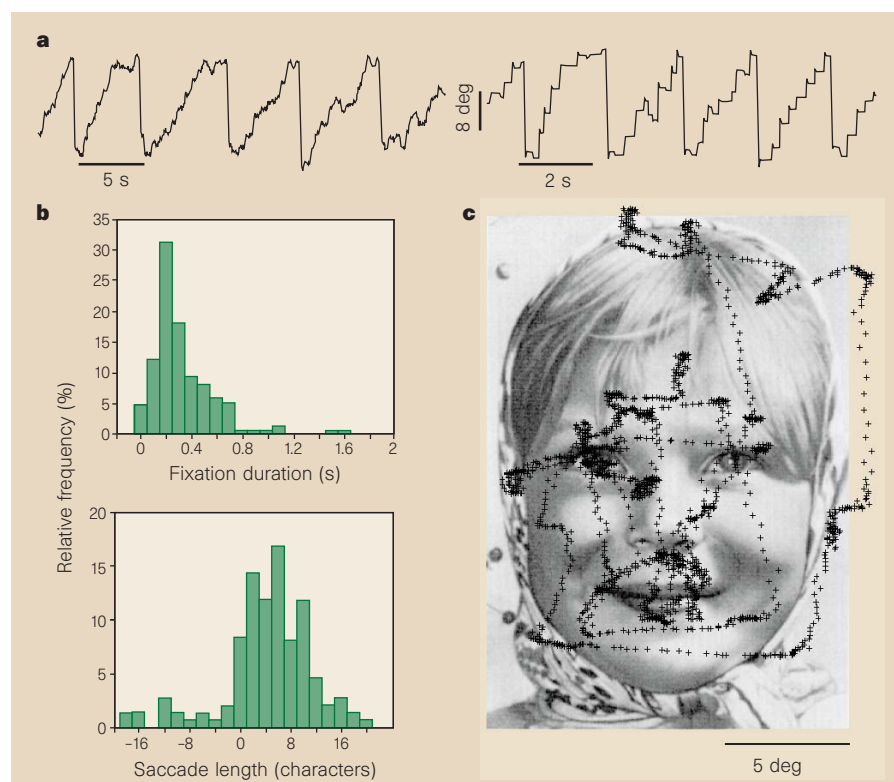


Figure 1 Recordings of head and eye movements. **a**, Head movements from AI (left) and eye movements from a control subject (right) during text reading recorded using a head-mounted search coil. Gaze position was sampled at 40 Hz. The horizontal eye or head displacement (left is down) is plotted against time. The slight overshoot in the eye-movement records is due to lens distortion⁷. AI is considerably slower overall and her head stability is not as good as that of the control subject. There seems to be a small, high-frequency tremor, probably reflecting the inherent instability of the head over the eye. **b**, The distribution of fixation durations and head movement sizes during single-sentence reading. Records were divided into fixation and movement periods on the basis of a velocity exceeding 10 deg s^{-1} over a 125-ms period. This eliminates the small-tremor head movements during fixation but allows for the detection of the onset of larger movements. AI's head movements are characteristic of normal eye-movement during reading. **c**, AI's head movements while viewing pictures, taken from the study of picture scanning by Yarbus⁴. The pictures were viewed for 20 s and head position was sampled at 100 Hz.

reading have a saccadic character and show many of the features that characterize eye movements.

Our subject, AI, is a 21-year-old female university undergraduate. As a result of an, apparently congenital, extraocular muscular fibrosis resulting in ophthalmoplegia, AI has had no eye movements since birth. However, she reports no major visual problem associated with her deficit and receives no extra assistance, either with reading or writing, in her studies.

The presence of an optokinetic nystagmus (OKN) response is often taken as an indication of the presence of any residual eye-movement function². We used a sensitive eye-movement recorder³ to track her eyes during fixation and when presented with a large-field sinusoidal grating. The only eye movements that we recorded were very restricted drift movements ($\pm 0.5 \text{ deg}$ at most) which were not linked to the stimulus motion. Normal subjects showed a standard OKN response to this stimulus which could not be suppressed.

Reading provides one of the clearest, and

best characterized¹ examples of how visual processes and eye movements are coordinated to acquire information. Eye movements during reading have a number of defining features. The text is scanned in a saccadic manner, the eyes alternating between short fast movements and fixations, where the eye is stable. Most saccades are to the right, although a small proportion are regressive. Rightwards saccades are typically between seven and nine characters long, although the length can vary considerably depending on the text and the individual reader. Steady fixation is maintained between saccades, typically lasting between 200 and 250 ms. It is during fixation that information is gathered from the text. At the end of each line, subjects make a large leftwards return saccade that orients the eyes to the beginning of the next line.

AI's overall reading speed for standard passages of text with a wide range of difficulty was $257 \text{ words min}^{-1}$ (range 183 to 435), which is consistent with slow but not abnormal reading¹. This impressive reading speed is supported by movements of her head (Fig.

1a). AI's head movements have a saccadic pattern; most are rightwards with a few regressive movements. The forward fast movements have a modal size of six characters, and fixations have a normal average duration of 200 ms (Fig. 1b). In addition, AI makes the characteristic large return movement at the end of each line.

The saccade/fixate characteristic of eye movements is not restricted to reading, but is used in a range of visual scanning situations⁴ such as viewing a picture (Fig. 1c). Like Yarbus's original subjects, AI orients to locations of interest in the picture in a saccade-like manner, showing that her saccadic head movements are not restricted to reading and seem to be her general orienting mechanism.

Although AI's deficit is a peripheral one, her case suggests that saccadic movements, of the head or the eye, form the optimal sampling method for the brain. Given the additional mass of the head, and the non-specialization of the head in humans to make small saccadic movements (although such movements are common among other species⁵), the fact that such saccadic movements are adopted as an adaptive strategy by this subject indicates that the costs of adopting a new sampling strategy, such as smoothly scanning the display, outweigh the costs of moving the whole head in a saccade-like fashion many thousands of times a day.

When making large orienting movements, unimpaired subjects make combined head and eye movements. Recent physiological studies of these combined head and eye movements have focused on the role of the superior colliculus⁶, a midbrain region known to be important in visually guided motor action. Stimulation of neurons in the intermediate and deep layers of the superior colliculus of the rhesus macaque can lead to combined head and eye movements, demonstrating a projection from the superior colliculus to the head control centre in the brainstem. Such results suggest that an adaptation of neural function in the superior colliculus could be responsible for the total transfer of saccadic movements from the eyes to the head shown by this subject.

Iain D. Gilchrist, Valerie Brown
John M. Findlay

Centre for Vision and Visual Cognition,
Department of Psychology, Durham University,
Durham DH1 3LE, UK
e-mail: j.m.findlay@durham.ac.uk

Site of particle selection in a bivalve mollusc

Bivalve molluscs form dense populations that exert profound effects on the particle loads and phytoplankton composition of coastal waters¹. It has long been known that bivalves can select among different particle types, including selecting against those of poor nutritional value²⁻⁵, but because of difficulties in observing particle transport processes in the pallial cavity *in vivo*, the mechanism of selection was not known. We now use a combination of video endoscopy⁶ and flow cytometry⁷ to show that oysters can select living particles from non-living detritus on the gills. Our methods could aid the study of suspension feeding in many animal groups.

Oysters dominate estuarine bivalve assemblages throughout the world and form dense reefs that may strongly affect the seston of estuaries⁸. To determine whether oysters can select among the mixture of living and non-living particles, we

studied the western North Atlantic (*Crassostrea virginica*) and the Pacific (*C. gigas*) oysters. We observed particle transport directly using a surgical endoscope, and with the aid of a micromanipulator we positioned a sampling pipette to sample particles from ciliated transport tracts. We distinguished and counted particle types using flow cytometry.

We fed the oysters a mixture of two equally sized but qualitatively different particle types, the red-coloured microalga *Rhodomonas lens*, and ground dead leaves and stems of the cord grass *Spartina alterniflora*. The *S. alterniflora* had been lying above the highest extent of the tide for at least four months and presumably was less nutritious than the living microalgal cells. *Spartina* spp. salt marshes line the fringes of most eastern North American estuaries and *Spartina* spp. detritus is an important component of the seston⁹.

Oysters have a plicated, heterorhabdic gill with principal and ordinary filaments that allow transport of particles dorsally or ventrally on the gill. Particles may be moved dorsally on the gill to a ciliated tract that drives particles anteriorly in a slurry. Alternatively,

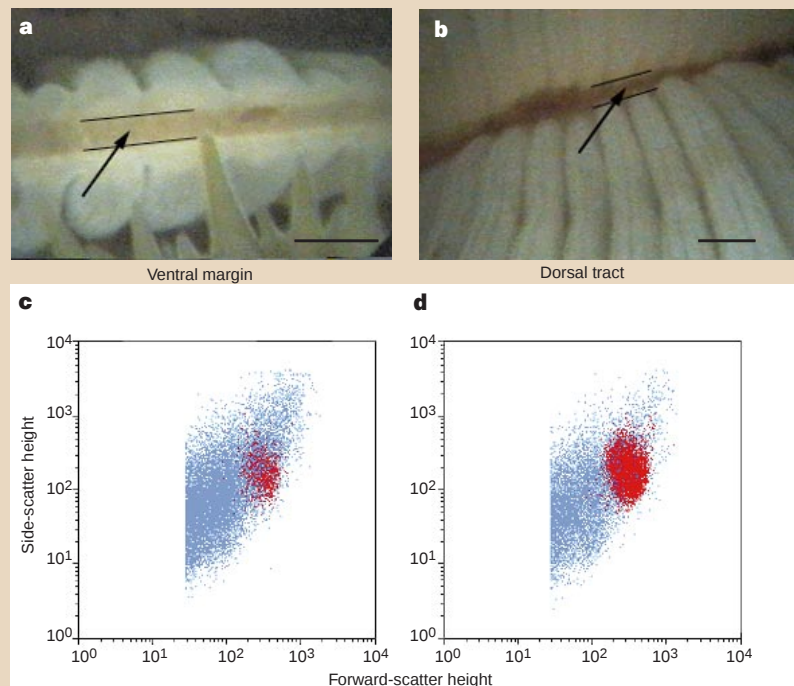


Figure 1 Feeding selectivity by oysters. **a**, Frame from video of the ventral ciliated tract of *C. gigas*, showing concentration of beige *S. alterniflora* detrital particles. **b**, Dorsal ciliated tract, showing concentration of red *R. lens* algal particles. **c**, Flow cytometric plot for *C. virginica* showing relative abundances of *R. lens* (red) and *S. alterniflora* detrital particles (blue) in the ventral margin, and **d**, from the dorsal tract. *C. virginica* (Friday Harbor Laboratories) and *C. gigas* (Southampton College Marine Laboratory) were exposed in a static chamber to 10^4 – 10^6 cells per litre, roughly equally distributed between live *R. lens* and detrital *S. alterniflora* particles. Particles were sampled after equilibration (30 min). Pseudofaeces were sampled and disaggregated for analysis. Particles were analysed by flow cytometry (Becton Dickinson FACScan bench-top flow cytometer) using detectors for forward scatter, side scatter, chlorophyll (fluorescence >650 nm), and phycoerythrin (fluorescence 560–590 nm). Heterogeneity among sorting efficiencies was significant (Kruskal-Wallis test, $P < 0.05$), showing positive selection for *R. lens* in the dorsal tract and enrichment of detrital particles in the ventral tract and pseudofaeces.

1. Rayner, K. & Pollatsek, A. in *Attention and Performance XIII* (ed. Coltheart, M.) 327–362 (Erlbaum, London, 1987).
2. Carpenter, R. H. S. *Movements of the eyes* (Pion, London, 1980).
3. Crane, J. D. & Steele, C. M. *Applied Optics* 24, 527–537 (1985).
4. Yarbus, A. L. *Eye Movements and Vision* (Plenum, New York, 1967).
5. Land, M. F. in *Eye Movement Research* (eds Findlay, J. M., Walker, R. & Kentridge, R. W.) 63–76 (Elsevier, Amsterdam, 1995).
6. Freedman, E. G., Stanford, T. R. & Sparks, D. L. *J. Neurophysiol.* 76, 927–952 (1996).
7. Deubel, H. & Bridgeman, B. *Vision Res.* 35, 529–538 (1995).

particles may be moved ventrally towards a ciliated tract that transports particles anteriorly in a cohesive mucus string. Both tracts transfer particles to the labial palps¹⁰.

Although roughly equal amounts of *S. alterniflora* detritus and *R. lens* microalgal cells were fed to the oysters, the rejected material in the form of pseudofaeces was strongly enriched in detrital particles (Fig. 1), indicating strong selection in favour of living particles. Video endoscopy revealed a strong differential selective transport of material enriched with *R. lens* towards the dorsal tract and material enriched with dead *S. alterniflora* particles towards the ventral tract. The latter tract transferred particles to the labial palps, and then rejected them as pseudofaeces. By contrast, the contents of the dorsal tract were transferred to the palp for further processing. Direct sampling between the palp lamellae showed that little additional sorting occurred and presumably the unaffected slurry of *R. lens*-dominated particles was transferred to the mouth.

If *R. lens* or *Tetraselmis* sp. (a green microalga) particles were fed alone to the oysters, nearly all algae were transported to the dorsal ciliated tract. Similarly, if detrital *S. alterniflora* particles were fed alone, the oysters seemed to transfer most of the material to the ventral tract, eventually to be rejected. This would suggest that the decision to accept or reject particles is not contingent on the presentation of a choice to the bivalves. On the other hand, preliminary evidence suggests that after four hours of feeding on microalgae, *R. lens*-rich material is redirected towards the ventral tract. This suggests that gut fullness might influence rejection of even good particles after a period of time.

Our results show that oysters can differentiate between nutritious and detrital particles. The mechanism is not based on particle size, but may be related to chemical cues, perhaps related to properties of the particle surfaces. Because oysters are strong interactors with estuarine ecosystems, this ability to differentiate between particle types should have profound effects on the composition and level of the seston, especially where resuspension brings rejected pseudofaeces back to the water column. Endoscopic examination and determination of the locus of sorting provides a direct connection between individual function and processes at the ecosystem level.

J. Evan Ward

Department of Marine Sciences,
University of Connecticut, Groton,
Connecticut 06340, USA

Jeffrey S. Levinton

Department of Ecology and Evolution,
State University of New York,
Stony Brook,
New York 11794, USA
levinton@life.bio.sunysb.edu

Sandra E. Shumway

Natural Sciences Division,
Southampton College of Long Island University,
Southampton, New York 11986, USA

Terry Cucci

Bigelow Laboratory for Ocean Science,
West Boothbay Harbor, Maine 04575, USA

1. Dame, R. F. *Bivalve Filter Feeders in Estuarine and Coastal Ecosystem Processes*. (Springer, Berlin, 1993).
2. Iglesias, J. I. P., Navarro, E., Jorna, P. A. & Armentia, I. J. *Exp. Mar. Biol. Ecol.* **162**, 177–198 (1992).
3. Pastoureaud, A., Heral, M., Proust, J., Razet, D. & Russu, P. *Oceanol. Acta* **19**, 79–88 (1996).
4. Newell, R. I. E. & Jordan, S. J. *Mar. Ecol. Prog. Ser.* **13**, 47–53 (1983).
5. Newell, C. R., Shumway, S. E., Cucci, T. L. & Selvin, R. J. *Shellfish Res.* **8**, 187–196 (1989).
6. Ward, J. E., MacDonald, B. A., Thompson, R. J. & Beninger, P. G. *Limnol. Oceanogr.* **38**, 265–272 (1993).
7. Shumway, S. E., Cucci, T. L., Newell, R. C. & Yentsch, C. M. J. *Exp. Mar. Biol. Ecol.* **91**, 77–92 (1985).
8. Kennedy, V. S. & Newell, R. I. E. *The Eastern Oyster: Crassostrea virginica*. (Maryland Sea Grant College, College Park, 1996).
9. Newell, R. I. E. & Langdon, C. J. *Mar. Ecol. Prog. Ser.* **34**, 105–115 (1986).
10. Ward, J. E. *et al. Biol. Bull.* **186**, (1994).

In-store music affects product choice

Royalty payments for non-broadcast commercial uses of music in 1995 amounted to £53.8 million in the UK alone¹. Research on music and consumer behaviour^{2,3} has, however, almost completely ignored the potential effect of in-store music on purchasing and particularly on product choice. By investigating the purchasing of German and French wines we have found that musical 'fit' has a profound influence on product choice.

Specific musical pieces may activate superordinate knowledge structures⁴, suggesting how in-store music could influence product choice. For example, stereotypical French music should activate superordinate knowledge structures concerning France, so

Table 1 Summary of results by type of music

	When French music played	When German music played
Bottles of French wine sold	40 (76.9%)	12 (23.1%)
Bottles of German wine sold	8 (26.7%)	22 (73.3%)
Extent to which music made respondents think of France (0 = not at all, 10 = very much)	6.25 ± 3.34	2.5 ± 3.68
Extent to which music made respondents think of Germany (0 = not at all, 10 = very much)	1.52 ± 2.08	6.08 ± 3.73
Usual wine preference of respondents (0 = French, 10 = German)	3.54 ± 3.44	5.58 ± 2.78
Ratings from the questionnaires were on a scale of 0 to 10. Mean ratings ± s.d. are given.		

priming the selection of products such as French wines. Similarly, stereotypical German music should activate related knowledge and prime the selection of products such as German wines.

To test this, four French and four German wines were displayed in a supermarket drinks section. The wines were matched between the countries for their price and dryness or sweetness. Each of the four shelves contained one French and one German wine and appropriate national flags. The position of the wines on the shelves was changed halfway through the two-week testing period. French accordion and German Bierkeller pieces were played on alternate days from a tape deck situated on the top shelf. Shoppers buying wines from the display were asked to complete a questionnaire by two experimenters posing as customers, with 44 shoppers (54%) consenting and the rest typically offering constraints on available time as a reason for declining.

French wine outsold German wine when French music was played, whereas German wine outsold French wine when German music was played ($P < 0.001$) despite an overall bias in favour of French over German wine sales (Table 1). Questionnaire responses indicated that the French music made respondents think of France and the German music made them think of Germany ($P < 0.001$). Respondents did not differ in their general preference for wines from these two countries ($P > 0.05$), and only six respondents answered 'yes' to the question, 'Did the type of music playing influence your choice of wine?'.

Customers did not seem aware of the effect that music had on their selections. Given recent controversy over subliminal perception⁵ it would be interesting to discover whether they were really as unaware of the effects of musical 'fit' as their questionnaire responses suggested. Future research could investigate the effects of music relative to silence, or relative to the effects of music from a country that does not produce wine; the mediating effect of individual differences⁶; and whether musical 'fit' has more influence on product choice when customers are undecided⁷.

Adrian C. North, David J. Hargreaves

Jennifer McKendrick

Music Research Group, Department of Psychology,
University of Leicester, University Road,
Leicester LE1 7RH, UK

1. *The Value of Music* (Natl Music Council, London, 1996).
2. North, A. C. & Hargreaves, D. J. in *The Social Psychology of Music* (eds Hargreaves, D. J. & North, A. C.) 268–289 (Oxford Univ. Press, 1997).
3. Areni, C. S. & Kim, D. *Adv. Consumer Res.* **20**, 336–340 (1993).
4. Martindale, C. & Moore, K. J. *Exp. Psychol. Hum. Percept. Perform.* **14**, 661–670 (1988).
5. Theus, K. T. *Psychol. Marketing* **11**, 271–290 (1994).
6. Kemp, A. E. in *The Social Psychology of Music* (eds Hargreaves, D. J. & North, A. C.) 25–45 (Oxford Univ. Press, 1997).
7. Bodenhausen, G. V. & Wyer, R. S. J. *Personal. Social Psychol.* **48**, 267–282 (1985).

The illusion of choice in patient power

Demanding Medical Excellence: Doctors and Accountability in the Information Age

by Michael L. Millenson

University of Chicago Press: 1997. Pp. 416.

\$24.95, £19.95

George Annas

Physicians and journalists love to tell stories about individual patients. Taking these tales seriously is the heart of medicine as art — and the Achilles' heel of medicine as science. As more clinical trials are published, more meta-analyses done and more data collected on outcomes and clinical pathways, medicine has the possibility of becoming more scientific (as well as more effective and safer). But can science-based medicine be introduced in a way that cuts costs, increases safety and retains the essence of medicine — that is, the doctor–patient relationship within which treatment decisions about (unique) individuals are made?

Michael L. Millenson, a journalist and former reporter for the *Chicago Tribune*, is optimistic about this question. After traveling around the United States and talking with 'information revolution' gurus such as David Eddy, John Wennberg and David Blumenthal, he thinks the answer is yes.

Millenson makes his case by weaving together descriptions of his visits to places (such as the Mayo Clinic) engaged in infor-

mation-based programmes of quality improvement with stories about individual patients. Take Ray Spackman, a 63-year-old dairy farmer with coronary artery disease. His physicians use a computer program to calculate his likely five-year survival rates among treatment alternatives: medication alone, 10 per cent; angioplasty, 26 per cent; and bypass surgery, 50 per cent. Perhaps, needless to say, Spackman and his surgeons decide on bypass surgery. Or consider 32-year-old Lisa Ferris, a member of a health maintenance organization who is informed by her nurse-midwife health-care provider that "she doesn't need to return for a Pap smear for two or even three years" because the organization has changed its screening rules. The aim is to take the money saved by reducing the frequency of screening "low risk women like Lisa Ferris" and spend it on outreach programmes to increase the number of higher risk women who get screened.

Millenson seems to believe that these stories prove that more information always provides patients with more power. But these anecdotes illustrate that this is not necessarily true. Spackman is getting important information, but he really has little choice on the basis of it. He may also want to know how the outcomes of his surgeons compare with those of others in the region, and even in the country. But even this additional information will be useless to him if he has no choice

of surgeons in his health plan. And what is Ferris to do with her new information? She is simply told of a change of screening policy made without her input and apparently leaving her with no choice but to accept it.

I, like Millenson, want to believe that giving patients access to more information can improve the quality of their care. But providing it is more problematic than it seems. First, most of the new information about outcomes, pathways and physicians is available to hospitals, health-care plans and large employer purchasers, but not to patients. That employers can use this information to decide what health plans to purchase for their employees is convenient for them, but it hardly empowers employees who are stuck with the health plan 'choice' made by their employers. Only when patients choose their own health plans can this information empower them.

Second, what new information is available has seldom itself been subjected to clinical trials, so its validity and reliability remains suspect. One cannot make a credible case against anecdotal medicine by telling anecdotes.

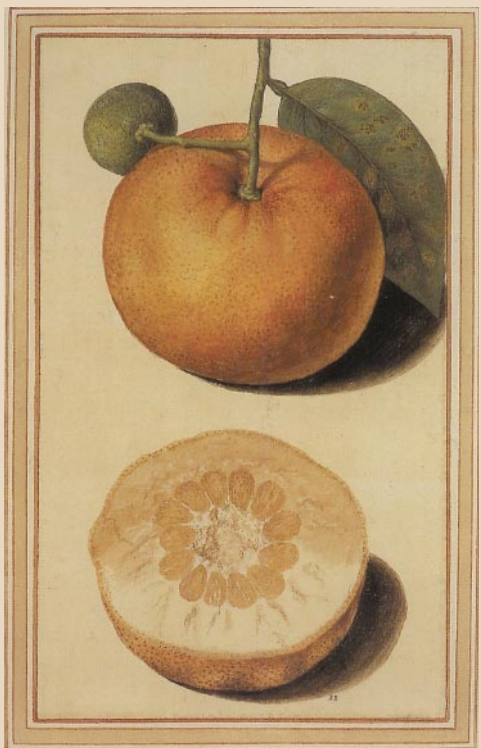
Third, quality is much harder to measure than Millenson or his informants think, especially if one believes that it is important to include quality of life as seen from the patient's perspective. It is easy to count the number of Pap smears performed and record mortality rates, but one is a process and the other a measure of outcome, and we must know how they relate to each other to make any meaningful assessment of treatment quality — and even then we will know nothing about quality of life. The goal of quality medicine is laudable (as are the two often competing goals of universal access to care and cost containment). But, until we agree on how quality is to be defined and by whom, we will not know what information to collect, analyse and disseminate.

Like many other information-age prophets, Millenson adopts the analogy of the factory for the hospital and sees no reason why factory methods of quality improvement cannot work in hospitals (and some probably can). The problem with the factory model is that patients are not products: each is an individual, and it is inherently wrong to treat them like fungible inanimate objects. The problem with the closely related market model is that even providing all the information in the world to the individual patient will not turn that patient into a 'consumer' or a customer.

Increasing patients' rights seems a more appropriate way to empower patients than giving them more information. Millenson should have spent more time acknowledging that it was judges, not physicians or health

Of pith and kin

Vincenzo Leonardi's drawings earned lavish praise from Giovanni Battista Ferrari, compiler of *Hesperides*, an ambitious attempt at a complete classification of citrus plants, published in Rome in 1646. "O Vincenzo, you double nature with your art, since you produce real fruit by what you paint within this volume... Is it not another miracle, that the fruits which are enjoyed here need not be torn from their branches? They are solely for the pleasure of the eyes. It is forbidden to pluck them with the hand, to eat them with the mouth; and so these marvellous fruit may be preserved for eternity." The antiquary and natural scientist Cassiano dal Pozzo commissioned much of the work for *Hesperides* and amassed an enormous collection of drawings and engravings of antiquities and scientific subjects. *The Paper Museum of Cassiano dal Pozzo: Citrus Fruit* by David Freedberg and Enrico Baldini (Harvey Miller, £150; distributed in North America by University of Toronto Press, \$225) is one of a series of books produced by delving into this material.



plans, who imposed information-sharing on medicine through the doctrine of informed consent. Furthermore, legal mechanisms, including licensing and malpractice litigation, currently provide the only quality standards (inadequate though they are) in medicine. There is a lack of enthusiasm for sharing even minimal quality information with patients, and such things as medical malpractice lawsuits and disciplinary proceedings should teach us that physicians are unlikely to share performance experience with patients without a law that compels this disclosure.

Millenson's book is a well-written and easily digestible entry into the world of the health information prophets. But if you are looking for something that explores what information matters to physicians and patients, and how it can be collected and used in a way likely to benefit patients by providing them with better and more personal care, you won't find it here. We must have some agreement on the goals of medicine before we know what to count to help us to decide if 'we' are doing 'better'. Only then can we tell if we are in the midst of an information 'revolution' or just getting more information faster in ways that have little impact on our medical treatment decisions. □

George Annas is in the Health Law Department, Boston University School of Public Health, 715 Albany Street, Boston, Massachusetts 02118, USA.

Chemist as catalyst

Justus von Liebig: The Chemical Gatekeeper

by William H. Brock
Cambridge University Press: 1997. Pp. 374.
£50, \$79.95

Kostas Gavroglu

Justus von Liebig (1803–73) was one of the most intriguing figures in the history of chemistry, and his personality and career provide ample material for exploring many themes in the history of science. Appointed to a little-known university, he transformed chemistry teaching and created the prototype of a modern research school that made the University of Giessen in Germany famous the world over.

Liebig was almost obsessively preoccupied with convincing others of the catalytic role of chemistry, with the aim of giving scientific status to pharmacy, physiology, agriculture and husbandry; this in turn changed the way in which pharmacists, physiologists, doctors and botanists thought about and practised their disciplines. He was also an assiduous popularizer, and his *Familiar Letters on Chemistry*, written over a 20-year period, was hugely successful in informing the public of developments in chemistry not just in Germany but throughout Europe and

the United States.

For a historian of science, Liebig's career encapsulates the tension between the private and the public scientist, helps to probe the relationship between the necessary precision of scientific language and the rhetoric of public discourse, encourages contemplation of the difficult issue of popularization, reflects the symbiosis of teaching and research, provides a well-defined laboratory space to examine the legitimization of new practices, and draws attention to the complicated links between the social repercussions of public disputes and their content.

William H. Brock, a professor of history of science at the University of Leicester in England and one of the foremost historians of chemistry, has written an admirable biography of Liebig that gives most of these issues a balanced yet rigorous treatment. Although Brock emphasizes Liebig's British connections, he provides an international perspective by portraying Liebig as a gatekeeper: "In adopting the gatekeeper image, I have in mind the ways in which Liebig acted as an entrepreneur and propagandist for the extension of chemistry's boundaries."

Liebig's public agenda was intertwined with a quasi-methodological programme to place chemistry at the core of all activities that would further the well-being of nations. Chemistry, for Liebig, was a mature science, and there were other disciplines that, through the intervention of chemistry, could themselves acquire scientific status. They would consequently be more effective and give rise to all kinds of new fields.

Liebig's involvement in modernizing the training of pharmacists began in the 1830s, when Philipp Lorenz Geiger decided to upgrade *Annalen der Pharmacie*. Time and again, both men stipulated that pharmacists needed to learn more about the chemical constitution of drugs, and that medical education had to include pharmacy and chemistry as a significant part of its curriculum.

A few years later, in 1840, Liebig had completed his *Agricultural Chemistry*, arguing that agriculture had to transcend empirically systematized techniques and transform itself into a science through the extensive help of chemistry. Liebig also had a lot to say in the book about plant physiology; and he initiated lines of argument that he would further develop in his *Animal Chemistry* published in 1842, where he expanded his thesis about the interrelationship of animal physiology and animal pathology.

Academic controversies and public disagreements are almost synonymous with the careers of high-profile scientists. Liebig was no exception, and he was often hot-tempered, spiteful and insidious. His dealings with people involved in disputes were further strained by his unwillingness to admit that many of his suggestions and techniques in agricultural chemistry and hus-



Liebig: "hot-tempered, spiteful and insidious... but a master at realizing when a quarrel might get out of hand".

bandry were neither well defined nor capable of being corroborated with precision.

But, as is often the case with those whose main preoccupation is the advancement of their ideological agendas rather than the relentless pursuit of their adversaries, Liebig was a master at realizing when a quarrel might get out of hand and was then quick to bring it to an end. Interestingly, he avoided the long-standing controversies of organic chemistry, especially those relating to atomic weights and chemical structure.

Liebig forged a special relationship with many British chemists, industrialists and politicians and even with some members of the royal family. Most importantly, through this relationship, radical changes were brought about in Britain's scientific education. The translation of his books, his forceful propaganda, his well-placed connections and the single-mindedness with which he pursued even several unconfirmed theories and techniques led to the transformation of English farming and the practice of agricultural chemistry.

In 1845, Liebig was offered the chance to settle in London. Many in Britain had hoped that he would accept the professorship at the newly established Royal College of Chemistry, so that they could reap all the teaching and research innovations that he was successfully developing at his own university in Giessen.

Liebig declined the offer and instead recommended Wilhelm August Hoffmann, one of his former students, who stayed in the post for 27 years. Hoffmann, together with the ever-increasing number of British students who spent time at Liebig's laboratory in Giessen and returned to Britain to hold various posts, became the most effective publicists for Liebig's many agendas.

Liebig's frequent trips to Britain and his widespread contacts made him an acute

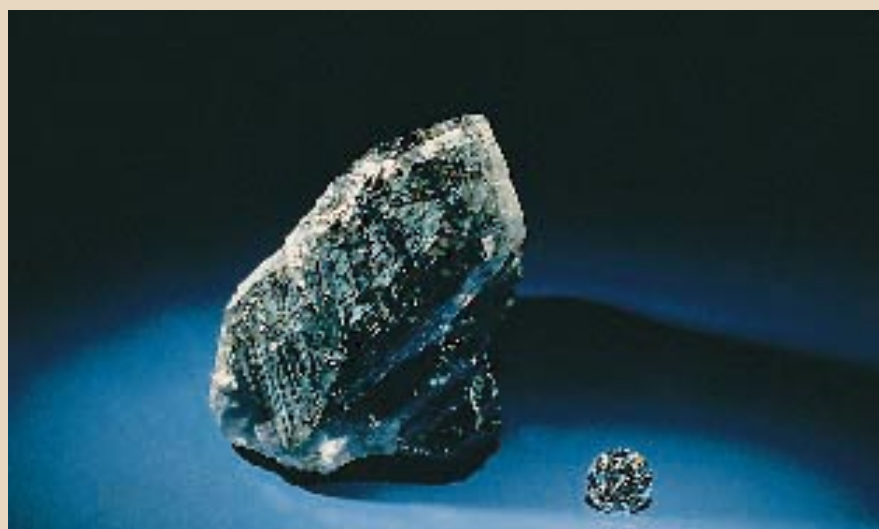
observer of the ways of the British, as shown in a letter he wrote in 1844 to Michael Faraday: "What struck me most in England was the perception that only those works of a practical tendency awake attention and command respect; while the purely scientific, which possess far greater merit, are almost unknown." Liebig could not have expressed better the difference in style and approach between British scientists and those on the continent.

There have been many excellent articles on various aspects of Liebig, especially the research school he created at Giessen. Brock's excellent biography is, however, the first published in English since 1901. It includes the translation of the report to the Prussian government on the chemical laboratory at Giessen by C. W. Bergemann, a professor of pharmacy at the University of Bonn, as well

as a valuable list of Liebig's British and American students.

By using the gatekeeper metaphor, Brock has satisfied a pressing need for a work that would systematically present Liebig's multifarious activities. It is worth noting Brock's succinct expression of Liebig's own perception of himself as a gatekeeper: "He was the voice that was in position to draw attention to the inexhaustible supplies of neglected wealth that the commercial and industrialized nations of the world could mine by following the chemical road... and how chemists themselves saw their roles intellectually as possessing civic worth in a modern society."

Kostas Gavroglu is in the Department of the History and Philosophy of Science, University of Athens, 37 John Kennedy Street, 161-21 Athens, Greece.



Diamonds in spades

Some three billion years ago, in the Earth's mantle, the hardest natural substance in the world was formed and stored. Kimberlite and lamproite magmas later brought it to the surface. Now researchers use it for creating high-pressure cells to investigate planetary interiors and dense matter, mimic the Earth's core or produce solid hydrogen.

The scientific story of the diamond is one facet of a new exhibition at the American Museum of Natural History in New York, "The Nature of Diamonds," which runs until 26 April 1998. The exhibition is accompanied by a book of the same name written by the museum's curator of gems and minerals, George E. Harlow (Cambridge University Press, \$75, £55 (hbk); \$29.95, £19.95 (pbk)).

Transliterated from the Greek 'adamao' meaning 'I tame' or 'I subdue', the stone's name has become connected with power and strength, and these associations are explored in cultural and historical facets of the exhibition. These sections lead on to a stunning array of jewels from the Middle Ages to the twentieth century,

including the Incomparable (pictured), weighing 407.48 carats — the third-largest cut diamond in the world.

But the exhibition offers plenty for interested scientists to get their teeth into. Interactive displays illustrate the main properties of diamond, such as its hardness, refractive index, conductivity and colour — the blue diamond, for example, gets its colour from minute amounts of boron.

The way diamonds are formed is explained in detail, with displays taking the observer through their history.

The use of diamond in research is also highlighted. As well as the diamond anvil cell, whereby two stones are squeezed together to create pressures of up to 4.5 million atmospheres, diamond is also a universal cutting tool and, because of its apparent iciness — heat-conducting ability — is used in electronics to manage heat.

For both science and spectacle, this exhibition is a real gem.

Alison Mitchell is an assistant editor of Nature.

Signs of the times

Comets, Popular Culture, and the Birth of Modern Cosmology

by Sara Schechner Genuth
Princeton University Press: 1997. Pp. 365.
\$49.50, £32.50

Donald Yeomans

My personal collection of comet memorabilia contains broadsheets, or flyers, issued for the appearances of comets Kohoutek and Halley around Christmas 1973 and in March 1986 respectively, and comet Shoemaker-Levy 9's impact with Jupiter in July 1994. The Kohoutek broadsheet asks "What will the Christmas monster bring?" The Halley flyer screams out "Behold! Halley's comet is coming! The end of mankind is near!" The broadsheet for Shoemaker-Levy warns that the comet's impact with Jupiter is surely an ultimatum from Almighty God: we must ban all crime, indecency and violence, or face global extinction from an Earth-hitting asteroid.

I am struck by the similarities in superstition between these recent broadsheets and those issued two or three centuries ago. The ancient fear of comets as signs or agents of terrestrial disaster has at least partly survived to the present day.

One of Sara Schechner Genuth's main points is that comets have been considered as such signs, or agents, of change throughout most of recorded history. Before the mid-seventeenth century, they were often viewed as atmospheric signs sent to warn sinful humanity that God was not at all pleased with its conduct. Even the demonstration, by Isaac Newton and Edmond Halley in the late seventeenth century, that comets were celestial objects orbiting the Sun did not sweep away the superstitious fear of comets. For Halley had noted that comets travel around the Sun in elongated orbits, so they could, from time to time, collide with Earth.

Thereafter comets were feared not as signs from an angry God but as agents of God's wrath. Their terrestrial effects were no longer seen as localized but rather as global, and Newtonian mechanics could be used to predict which comets might pose a threat. In the late seventeenth century, Newton invoked comets as vehicles for replenishing the planetary fluids consumed by the process of vegetation and putrefaction and for re-supplying the Sun and other stars with fuel so that these celestial fires could continue. For his part, Halley invoked cometary close approaches or collisions to explain the biblical flood, the periodic extinction and renewal of Earth's life forms, and even the origin of the world and its eventual demise.

Cosmologists in the eighteenth century expanded on these views. Comets could create, destroy, reform or restore planetary systems. But in the nineteenth century came



Woven warning: detail from the Bayeux tapestry showing Harold I being told of comet Halley. The eleventh- or twelfth-century tapestry in Bayeux, France depicts the Norman conquest of England.

Charles Darwin and, perhaps influenced by his views, which favoured slow evolutionary processes for explaining the development of life, the catastrophism of Newton and Halley was all but ignored until fairly recently.

In the past few decades, comet and asteroid collisions have once again been recognized as important processes in rearranging the surfaces of the planets and their satellites and abruptly changing the evolution of life on Earth. There is a growing consensus in the astronomical community that comet collisions with Earth may have laid down much of the thin layer of carbon-based molecules and water that allowed the formation of life 3,500 million years ago.

Subsequent cometary collisions may have caused mass extinctions, allowing only the most adaptable species to evolve further. We mammals may owe our pre-eminent position on Earth to a series of cometary collisions that eliminated our stronger, but less adaptable, competition — including the dinosaurs.

This book is not a general history of comets. Rather than focusing on the development of ideas about their motions or physical characteristics, the author is concerned primarily with the perception of comets throughout history. It is a scholarly, well-illustrated and accurate work. Nearly half the volume is devoted to footnotes and references, however, which leaves the reader with the annoying task of continually having to leaf back and forth between the two halves.

European and English perceptions of comets during the seventeenth and eighteenth centuries are emphasized. Little mention is made of other cultures — that of China, for example — that actively observed and recorded comets. But within the confines of the subject area presented, the author does a fine job. The book should make an important contribution to the history of astronomy. □

Donald Yeomans is at the Jet Propulsion Laboratory, Pasadena, California 91109, USA.

Getting down to business

The Scientist as Consultant: Building New Career Opportunities

by Carl J. Sindermann and Thomas K.

Sawyer

Plenum: 1997. Pp. 341. \$29.95, £18.15

William Bains

'Consulting' is famously identified with someone who steals your watch to tell you the time and, in this era of 'downsizing', has become almost synonymous with 'unemployed'. An exception to both calumnies is the technical consultant, especially the scientist. These are people who provide their expertise to many clients for a fee. But this book is not aimed at them, but rather at people who wonder whether consultancy is a career for them. Whether the career is right for you or not, the book is an excellent guide to help you decide.

What is a scientific consultant? The authors wrestle with this, and arrive at "a technically trained entrepreneur who makes available for a stated price his expertise, data, data analysis, evaluation and recommendations relevant to a client's needs", a catch-all they admit is unsatisfactory. They also emphasize the need for professional ethics as a consultant, citing a quaintly pre-1980s definition of a professional as one "who maintains a loyalty to a code of ethics that transcends his or her loyalty to the rest of the organisation" or to themselves, a tenet which, if adhered to by major consultancies, would send several of them bankrupt.

But above all the consultant is a business person, and consulting is a business. Consultants must be interested in the processes of both business and science. This means accepting the value of lawyers and accountants as advisers, sending off bills promptly and harassing clients who refuse to pay them, and 'selling'. Most scientists are unused to selling anything except ideas and, if you are not keen to try, then consultancy is not for you. Most consultancies fail, the authors believe, because of lack of aptitude for and interest in business. Squaring this with the Sisyphean task of keeping technically current requires real entrepreneurship. Scientific consulting is not just 'a job'.

The authors describe a rewarding career path from paid hobbyist to professional manager, which you can join or leave at will. They examine what sort of people might flourish in consultancy and why, how to escape from it, what the future is, and how people change, succeed or fail. They also give substantial detail on what consultants actually do. (The section on managing scientists is excellent — a 'must' for department heads as well as industrial managers.) The book is

stuffed with useful comments and guidance, including a very honest (if rather short) section on the downside of consulting. Consultants will enjoy putting names to the list of "clients from hell".

My only serious disagreement is with the authors' perception of big consultancy companies. Graduate entry to a large consultancy is not a viable route to a career in scientific consulting. Scientists are the drudges in such organizations, and do not rise to the top without radically altered goals; the leader of the 'science division' in one such consultancy publicly commented early in his post that research and development were a waste of money. Nor can they leave to set up on their own, as the competition clauses in their employment contracts will prevent them from competing as a consultant with their erstwhile employer. The route to scientific consultancy is clearly science first, consultancy later.

The book has a strong US bias, and 'rest of world' seems to mean not Europe but Africa. That said, non-American readers can easily sidestep the few parochial discussions.

This is a business book, because consultancy is a business. But, like science, the book is full of facts and hard detail, and does include the negative controls of business or scientific failure. It is an excellent guide to a fascinating career choice. □

William Bains is at Merlin Ventures,

Marquis House, 67/68 Jermyn Street, London SW1Y 6NY, UK.

e-mail: wbains@merlin-ventures.co.uk

Corrections

● Some incorrect star ratings were awarded to *The Colours of Life* by Lionel R. Milgrom in the review in *Nature* **389**, 687; 1997. They should have been "excellent" for range and style, "good" for depth, accuracy and accessibility, and "fair" for up-to-dateness. Our apologies for falsely raising expectations.

● In his review of *Molecular Systematics of Fishes* edited by T. D. Kocher and C. A. Stepien (*Nature* **389**, 30; 1997), John Long says that the book "only covers the largest living group of fishes, the teleostean fishes". The editors have asked us to point out that the book in fact contains a chapter on "Interrelationships of Lamniform sharks" — which are not, of course, teleosts.

● *Tropical Medicine and International Health*, reviewed in this year's New Journals supplement (*Nature* **389**, 145; 1997), resulted from the pooling of four previously existing journals, not three. The title not mentioned is the German *Tropical Medicine and International Health*, whose independent history extends back to 1897. *Antiviral Therapy*, reviewed on page 144 in the same supplement, has since its third issue been published by International Medical Press, not MediTech Media.

A new symmetrodont mammal from China and its implications for mammalian evolution

Yaoming Hu*, Yuanqing Wang*†, Zhexi Luo† & Chuankui Li*

* Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China

† Section of Vertebrate Paleontology, Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, Pennsylvania 15212 USA

A new symmetrodont mammal has been discovered in the Mesozoic era (Late Jurassic or Early Cretaceous period) of Liaoning Province, China. Archaic therian mammals, including symmetrodonts, are extinct relatives of the living marsupial and placental therians. However, these archaic therians have been mostly documented by fragmentary fossils. This new fossil taxon, represented by a nearly complete postcranial skeleton and a partial skull with dentition, is the best-preserved symmetrodont mammal yet discovered. It provides a new insight into the relationships of the major lineages of mammals and the evolution of the mammalian skeleton. Our analysis suggests that this new taxon represents a part of the early therian radiation before the divergence of living marsupials and placentals; that therians and multituberculates are more closely related to each other than either group is to other mammalian lineages; that archaic therians lacked the more parasagittal posture of the forelimb of most living therian mammals; and that archaic therians, such as symmetrodonts, retained the primitive feature of a finger-like promontorium (possibly with a straight cochlea) of the non-therian mammals. The fully coiled cochlea evolved later in more derived therian mammals, and is therefore convergent to the partially coiled cochlea of monotremes.

Systematic palaeontology

Class Mammalia

Subclass Theria

Order Symmetrodonta

Family Spalacotheriidae

Zhangheotherium quinquecuspidens gen. et sp. nov.

Holotype. IVPP V7466 (Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China), a well-preserved skeleton (Fig. 1) consisting of a partial skull, and most of the postcranial skeleton, including all cervical and thoracic vertebrae, forelimbs and pectoral girdle, as well as the excellent impressions of the hindlimbs and pelvic girdle, and ribs. Lumbar and sacral vertebrae are represented by poor impressions. Some tarsals and most caudal vertebrae are missing (for measurements, see Tables 1 and 2).

Etymology. Zhanghe, in honour of Zhang He, who collected and donated the holotype specimen to the Institute of Vertebrate Paleontology and Paleoanthropology; therium, beast (Greek); quin-

que, five (Latin); cuspis, point (Latin); dens, tooth (Latin), for the three main cusps plus two large accessory cuspules on the lower molars.

Locality. Jianshangou Valley (approximately 41° 41' 01" N, 120° 59' 30" E), about 32 km east of Chaoyang City, Liaoning Province, northeastern China¹.

Horizon. The Jianshangou Beds, consisting primarily of shales, are the lowest lacustrine intercalation in the neutro-basic volcanic beds of the Yixian Formation^{1–3}.

Associated fauna. The Jianshangou Beds have yielded diverse fossil fish⁴, the birds *Confuciusornis*, *Liaoningornis*^{5,6} and *Protarchaeopteryx*⁷, the theropod *Sinosauropteryx*⁸, and diverse gastropods, bivalves, ostracods, conchostracans and insects².

Age. The age of the Yixian Formation is equivocal. Vertebrate faunal correlation^{4,5} and previous radiometric dates⁴ suggest that the Jianshangou Beds are either of the latest Jurassic age, or near the Jurassic–Cretaceous transition⁶. An Early Cretaceous age was also suggested by invertebrate faunal correlation³, and supported by a

Table 1 Skeletal measurements of *Zhangheotherium quinquecuspidens*

Bone	Length (mm)
Dentary	30.2
Axis (C2)	5.0
Cervical vertebrae 3–7	10.0 (total length)
Thoracic vertebra 8	2.9
Caudal vertebra 2	3.5
Clavicle	11.6
Scapula (posterior border to glenoid)	17.2
Humerus	22.1
Ulna	21.5
Radius	17.1
Metacarpal III	6.5
Ilium	20.1
Ischium	8.5
Epipubis	9.1
Femur	22.0
Tibia	23.5
Fibula	23.5
Calcaneum	6.6
Metatarsal III	8.6

Table 2 Dental measurements of *Zhangheotherium quinquecuspidens*

Tooth	Length (mm)	Height (mm)
Lower dentition		
I1	1.2	
I2	0.5	
I3	0.5	
C	0.8	
P1	1.2	
P2	1.3	
M1	1.5	1.0
M2	1.6	(1.5)
M3	1.8	1.7
M4	1.7	1.8
M5	1.5	1.8
M6	1.1	1.0
Upper dentition		
M2		1.5
M3	(1.6)	1.5
M4	(1.6)	1.6
M5	1.8	1.5

Parentheses indicate approximate estimates.



Figure 1 *Zhangheotherium quinquecuspidens* (IVPP V7463, holotype). Stereophotographs (**a**) and outline (**b**) of the skeleton in ventral view of the dorsoventrally compressed specimen (broken lines indicate the morphology preserved in impressions). Vertical scale bar in **b** represents 1 cm. Abbreviations: ac, acromion of scapula; c1–c7, cervical vertebrae 1 to 7; ca1, canine; ca2–ca4, caudal vertebrae 2 to 4 (caudal vertebrae are incomplete); cd, coracoid process of scapula; cl, clavicle; cm, calcaneum; co? coronoid fossa?; cp 1–9, carpals 1 to 9; csp, crista parotica of petrosal; ctp, caudal tympanic recess of petrosal; dn, dentary; ep, epipubis; er, epitympanic recess; fc, fenestra cochlearis (round window); fe, femur; ff?, facial nerve foramen?; fi, fibula; frs, foramen for ramus superior; fst, fossa for stapedial muscle; fv, fenestra vestibuli (oval window); g, glenoid of the scapula; gj, groove for jugal (on the squamosal); gl, glenoid fossa of squamosal; hu, humerus; ic, interclavicle; if, infraspinous fossa; il, ilium; is, ischium; i1–i3, incisors 1 to 3; ju, jugal; L1–L6, lumbar vertebrae 1 to 6 (impressions only); lm, lambdoid crest; mas, mastoid exposure of petrosal; mc, metacarpoid; mf,

mandibular foramen; mg, meckelian groove; mp1–mp5, metacarpals I–V; mt1–mt5, metacarpals I–V; mst, manubrium of sternum; mx, maxillary; oc, occipital condyle; of, obturator foramen of pectoral girdle; on, odontoid notch for dens of atlas; pc, procoracoid; pcd, condylar process of dentary; pcl, preclavicle (a homologue of procoracoid, *sensu* ref. 50); pcr, coronoid process of dentary; pf, pterygoid fossa; pgc, postglenoid crest; pgd, postglenoid depression; pmx, premaxillary; pp, paroccipital process of petrosal; ppe, paroccipital process of exoccipital (incomplete); pr, promontorium of petrosal; prs, prootic sinus canal; ptc, post-temporal canal; pts, post-tympanic sinus; p1 and p2, premolars 1 and 2; ra, radius; r1–r13, thoracic ribs 1 to 13 (posterior thoracic ribs preserved only in impressions); sc, scapula; sl?, fossa for splenius? on the dentary; sm, stylomastoid notch; sp, spine of scapula; sq, squamosal; ss, supraspinous fossa of scapula; stb1–stb6, sternebra 1 to sternebra 6; stl, embryonic sternal bands; sym, mandibular symphysis; s1–s4, sacral vertebrae 1 to 4 (represented mostly by impression); th, attachment site for tympanohyal; ti, tibia; ts, lateral tarsal spur of ankle; t1–t13, thoracic vertebrae 1 to 13; ul, ulna; x, xiphoid process of sternum.

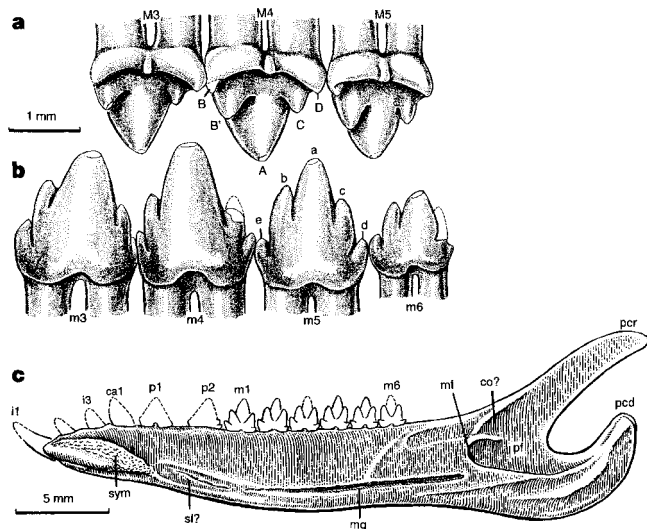


Figure 2 Dentition and mandible of *Zhangheotherium quinquecuspidens*. **a**, Upper molars M³⁻⁵; **b**, lower molars M₃₋₆ (left, labial views); **c**, mandible (right side, reconstruction in medial view). *Zhangheotherium* closely resembles *Spalacotherium* in lower molars, and *Peralesstes* in upper molars. *Spalacotherium* and *Peralesstes* were established on some dissociated lower and upper teeth, respectively, from the same fauna in the Late Jurassic of England¹⁰⁻¹². The associated upper and lower dentitions of *Zhangheotherium* suggest that *Spalacotherium* and *Peralesstes* are the lower and upper teeth of the same species, and that *Peralesstes* should be synonymized with *Spalacotherium*¹⁰⁻¹². Cusp B' is too large and positioned too far lingually to be the stylocone (cusp B). It could be either a homologue to the much smaller intermediate cusp in the comparable position in other symmetrodonts, or a neomorphic cusp. We follow ref. 45 for the designation of cusps A, B, C, D and E on the upper molars, and cusps a, b, c, d and e on the lower molars. For abbreviations, see Fig. 1 legend.

recent radiometric date⁹. According to our most recent field investigation, this date⁹ should be regarded as an upper age limit for the Jianshangou Beds.

Diagnosis. A spalacotheriid symmetrodont (Fig. 2) with the dental formula 3.1.2.5/3.1.2.6; differing from all other known symmetrodonts in having a hypertrophied cusp B' between the main cusp A and the stylocone (cusp B) (Fig. 2); distinguishable from other spalacotheriids^{10–12} in the conical shape of the main cusps and the lack of lingual and labial cingulids on the lower molars¹¹; distinctive from *Spalacotherium*¹¹ in having more robust and rounded main cusps that lack connecting cristae; unique among early mammals^{13–18} in having fused sternalbrae, and a posteriorly expanded xiphoid process.

Description and comparison

Zhangheotherium resembles other known spalacotheriid symmetrodonts^{10–12,19,20} in that the central cusp (a) and the two accessory cusps (b, d) of the lower molar form an acute triangle. The associated upper and lower teeth of *Zhangheotherium* could occlude into the embrasures of the opposing tooth row, a pattern unique to spalacotheriids among archaic therians¹⁰. This suggests that the teeth were used more for crushing and puncturing than shearing²⁰.

Zhangheotherium is unique among spalacotheriids in having short and weak cingula in the upper molars, and a relatively low trigonid and two large accessory cusps (cusps d and e) on the lower molars^{10–12,19,20}. Upper molars have a wide labial shelf and their cusp pattern resembles that of *Perales*, a Late Jurassic symmetrodont from England^{10–12}. Lower incisors are procumbent, with the first incisor I₁ being enlarged. C₁ is small and single-rooted, similar to I_{2–3}.

The dentary has a dorsally curved condylar process and a posteriorly tilted coronoid process. Like other symmetrodonts, it lacks an angular process (Fig. 2). The mandibular symphysis is oval, unfused and was probably mobile in life. The meckelian groove is narrow, and becomes shallow and faint anteriorly. The posterior part of the groove is separated from the mandibular foramen in the pterygoid fossa. The topographic relationships of these mandibular features (Fig. 2) are very similar to those of more derived therian mammals ('eupantotheres')^{10,16}, but differs from that of the earliest known therian *Kuehneotherium* (Late Triassic), and from those of the non-therian mammals *Morganucodon*²¹ and *Sinoconodon*²². A slightly rugose area along the anterior meckelian groove suggests the presence of the splenial, and a rough area near the base of the coronoid process is probably for a poorly developed coronoid, as occurs in the more derived therian *Henkelotherium* from the Upper Jurassic of Portugal¹⁶. No other postdentary bones are preserved.

The squamosal has an anteroposteriorly elongate glenoid fossa flanked by a low postglenoid crest (Fig. 3). The postglenoid region has a constricted neck between the glenoid and the broad cranial moiety that forms a wall lateral to the epitympanic recess. *Zhangheotherium* is similar in the latter two features to derived therians^{17,23,24}, but is different from non-therian mammals which have a narrow cranial moiety of the squamosal but no squamosal wall for the epitympanic recess^{25,26}. The postglenoid area of the squamosal has a posterolateral depression that resembles the broad external auditory meatus of *Vincelestes*¹⁷, the marsupial *Pucadelphys*²⁴ and placentals, in contrast to non-therian mammals in which the squamosal has no postglenoid region or the external auditory meatus^{21,22,27–29}.

The petrosal bones that contained the inner ear are complete but fractured and distorted in preservation. The promontorium that houses the cochlea has a cylindrical and finger-like shape (Fig. 3). Its structure is very similar to those of *Sinoconodon*²⁰, *morganucodontids*^{21,22,31}, *triconodonts*^{22,26} and *multituberculates*^{32–34}. As documented for a wide range of early mammals^{30,31,33–35}, a cylindrical and finger-like promontorium is closely correlated

with either a straight or slightly curved (but uncoiled) cochlea. We therefore infer from the slender promontorium of *Zhangheotherium* (Fig. 3) that it has a more or less straight (definitely uncoiled) cochlea. *Zhangheotherium* is very different from derived therian mammals, which have oval-shaped and more bulbous promontoria with cochleae coiled for at least 270 degrees^{17,23,33,36}.

The petrosal has a prominent paroccipital process that is similar to those of *morganucodontids*^{21,22}, *triconodonts*^{22,26}, *multituberculates* and *ornithorhynchids*³⁷. The post-temporal canal is positioned between the paroccipital region of the petrosal and lambdoidal crest of the squamosal, also a primitive character of mammals. *Zhangheotherium* has a large post-tympanic recess, and a broad and shallow epitympanic recess bound medially by a low crista parotica, and laterally by the squamosal. Both these features are characteristic of more derived therians.

Zhangheotherium has 7 cervical and 12 thoracic vertebrae. The postaxial cervical ribs remain unfused in adults, as in *morganucodontids*¹³ and *multituberculates*³⁸. All thoracic ribs are

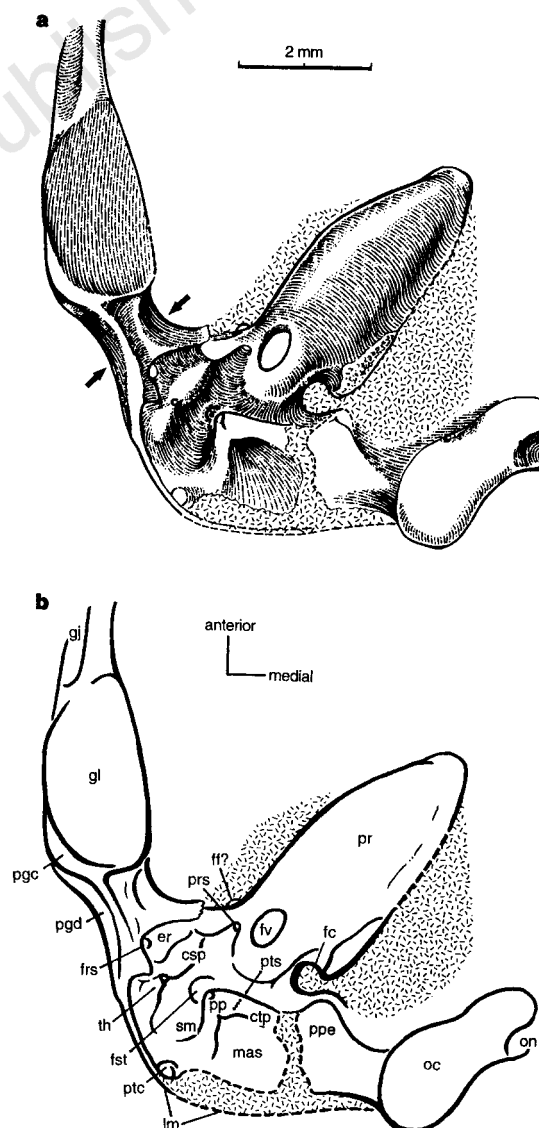


Figure 3 Reconstruction of the partial basicranium of *Zhangheotherium quiquecuspidens* (IVPP V7466). Right side, ventral view. Composite reconstruction (a) and outline with labels (b). Hatched pattern indicates the areas of damage or matrix. Arrows in a indicate the constricted neck of the squamosal. For abbreviations, see Fig. 1 legend.

robust and compressed anteroposteriorly. Six lumbar vertebrae and three or four sacral vertebrae are represented by impressions. The sacral vertebrae and three preserved caudal vertebrae have wide transverse processes (Fig. 1).

The interclavicle is V-shaped. It is clearly articulated with, and overlaps the ventral side of the anterior part of, the sternal manubrium. Its sharp posterior process is continuous with the sternal keel (Fig. 4). The manubrium, although fused to other sternbrae, is still discernible by its facet for the first thoracic rib. The short and broad lateral process of the interclavicle has a loose and mobile articulation with the clavicle. *Zhangheotherium* and multituberculates share strong similarities in the articulations of the interclavicle to the clavicle and to the sternum¹⁸. The lateral end of the clavicle has a spiral articular surface for the acromion of the scapula, allowing some mobility in the clavicle–scapular joint. The triangular scapular blade is demarcated from the glenoid by a scapular notch. The supraspinous fossa is developed along the entire length of the blade, but is much narrower than the infraspinous fossa. The spine is very high, with a prominent acromion. The coracoid process is hook-like. The glenoid faces ventrally, and is much narrower than the spherical humeral head. *Zhangheotherium* has mobile articulations of the interclavicle, clavicle and scapula that allow the clavicle to move and act as a pivotal strut³⁹ for a wider rotation of the scapula, as observed during the locomotion of the opossum⁴⁰.

The robust humerus of *Zhangheotherium* has a long deltopectoral crest. The intertubercular groove is narrow and deep, and separates the greater and lesser tubercles, the former being slightly wider than the latter. Torsion of the proximal end of humerus relative

to the distal end is about 30 degrees, about the same as in *Henkelotherium*¹⁶, and close to the 40 degree torsion in *Vincelestes*¹⁷. Distally, the humerus has an incipient trochlea for the ulna, a therian apomorphy^{16,17,41} that is absent in multituberculates^{14,18,42}. The humerus also has a weakly developed ulnar condyle, resembling those of non-therian mammals^{13,14,38,41}. The radial condyle is prominent and spherical, a primitive feature of non-therian mammals. Nine carpals are present. The pisiform is very large.

The pelvic bones are slender. The ilium is long and rod-like, and three times as long as the ischium. The acetabulum is partially preserved, showing a dorsal rim. Although the ischiopubic plate is not well preserved in IVPP V7466, its impressions indicate a shallow pelvis, differing from the deep pelvis of multituberculates³⁸. The epipubis is as long as the ischiopubic plate. As in multituberculates^{14,38}, the spherical femoral head is set off from the shaft by a well-defined neck, and the greater trochanter is directed dorsally. The calcaneum has an elongate tubercle and a well-defined peroneal tubercle. Its astragal process contacted the fibula, a plesiomorphy for therians²⁷. As in living monotremes and *Gobiconodon*¹⁵, *Zhangheotherium* has an external pedal spur. This spur is associated with a poisonous gland in the modern male platypus.

Phylogenetic implications for early mammals

The relationships of therians to extinct non-therian mammalian lineages have received much attention, as hypotheses of therian relationships are central to the understanding of early mammalian evolution^{27,43}. Therians have been hypothesized to be the sister taxon of either multituberculates^{18,26,27} or monotremes on the basis of some derived molar characters⁴⁴. Basicranial studies^{34,37} have suggested that monotremes and multituberculates were sister taxa to the exclusion of therians. There are few relatively complete fossils of archaic therians^{16,17}. The discovery of the skeleton of *Zhangheotherium* offers the first opportunity to assess the phylogenetic relationships of early therians using nearly all dental, basicranial and postcranial characters.

Symmetrodont mammals were previously known only from teeth and jaws. Owing to the lack of better evidence, the affinities of this lineage to the more derived therians have been based solely on dental evidence^{12,43–46}. It has been argued that dental characters are as homoplastic as non-dental characters^{27,43,47,48}, and the reliability of dental characters for inferring the relationships of major lineages of mammals has been questioned. *Zhangheotherium* has provided more extensive basicranial and postcranial evidence to corroborate the traditional hypothesis^{12,43–46} that symmetrodonts represent a part of the basal therian radiation. Our analysis (Fig. 5) suggests that *Zhangheotherium*, as the best-preserved taxon of all symmetrodonts, is a basal clade in therian phylogeny. It is more primitive than *Henkelotherium*¹⁶ and *Vincelestes*¹⁷ in retaining the interclavicle in its pectoral girdle. It also has a primitive dentition in which the lower molars have no distinct talonid, and the upper and lower molars occlude in the embrasures of the opposing tooth rows (Fig. 2), in contrast to more derived therians in which the lingual part of the upper molar occludes with the talonid of the lower molar^{16,45}.

The sister-taxon relationship between therians and multituberculates^{18,26,27} is strongly supported by the evidence from *Zhangheotherium*. Additional derived characters from *Zhangheotherium*, such as the clavicle–interclavicle joint and some features of the femur, corroborate the therian affinities of multituberculates (Fig. 5).

Implications for anatomical evolution

A large and reptile-like interclavicle is present in adult monotremes⁴⁹. In living marsupials (Fig. 4), it first appears embryonically as a separate ossification(s), but later becomes incorporated, at least in part, within the sternal manubrium in the adult⁵⁰. *Zhangheotherium* shows that the interclavicle was still present in at

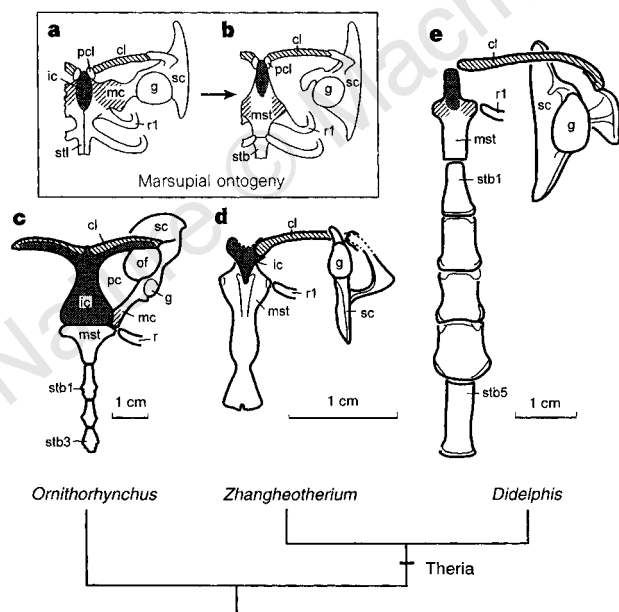


Figure 4 Comparison of the sternal apparatus and pectoral girdle of *Zhangheotherium* and living mammals. **a, b**, Embryonic (**a**) and adult (**b**) stages of the sternum and pectoral girdle in marsupials (modified from ref. 50). **c–e**, The adult sternal and girdle structures of *Ornithorhynchus* (modified from ref. 49) (**c**), *Zhangheotherium* (**d**) and *Didelphis* (CMNH c45, and several uncatalogued specimens at Carnegie Museum) (**e**). The chondral element of the interclavicle and the medial part(s) of the embryonic coraco–scapular plate are considered to be incorporated into the sternal manubrium in adult marsupials^{49,50}. *Zhangheotherium* has retained a separate interclavicle, a primitive character of non-therian mammals, but a more derived character of a mobile clavicle–interclavicle joint which is present in multituberculates and in a modified form in living therians. For abbreviations, see Fig. 1 legend.

least some archaic therians. The loss of the interclavicle, or its incorporation into the sternal manubrium, occurred only in more derived therians. The reduced interclavicle and the mobile clavicle–interclavicle articulation in *Zhangheotherium* are far more derived than the large interclavicle and the rigid clavicle–interclavicle articulation in living monotremes (Fig. 4).

In contrast to the sprawling posture of living monotremes, most living therian mammals have a more parasagittal posture (*sensu* ref. 42), with the elbows positioned close to the thorax^{39–41}. Such a forelimb posture has been hypothesized to have evolved only once in the common ancestry of multituberculates, living therians and their extinct relatives¹⁸, although this view has been contested⁴². Several osteological characters are considered to be crucial to the parasagittal posture of the forelimb of therians^{18,38,39,41}: a mobile joint between the clavicle and sternal apparatus (including the interclavicle); a greater tubercle much wider than the lesser tubercle; a narrow intertubercular groove⁴²; a humeral trochlea that constrains the movement of the ulna⁴¹; and possibly the lack of torsion of the humerus¹⁸ (but see ref. 42). The skeleton of *Zhangheotherium* provides additional information on the phylogenetic distribution of these characters.

Zhangheotherium, *Vincelestes*, modern therians, and multituberculates all have a mobile joint between the clavicle and the sternal apparatus. This allows the clavicle to function as a pivot for the shoulder joint, and allows a greater range of rotation of the scapula during locomotion^{39,40} than is possible in monotremes. However, torsion of the humerus and the large lesser tubercle relative to the greater tubercle in *Zhangheotherium*, *Henkelotherium* and *Vincelestes*¹⁷ suggest that the humeri of these archaic therians were more abducted than those of most living therians. The humerus of *Zhangheotherium* has an incipient trochlea but also retains a vestigial ulnar condyle. The trochlea is also incipient in the Late Jurassic *Henkelotherium*¹⁶, but more developed in the Early Cretaceous *Vincelestes*¹⁷. This differs from the fully functioning (but more primitive) ulnar condyle of multituberculates^{14,18,38,42}. The incipient trochlea in such archaic therians as *Zhangheotherium* and *Henkelotherium* represents an intermediate condition towards the less abducted posture of the forelimb in living therians.

These new data suggest that the mobility of the clavicle and scapula has a more ancient origin than the more parasagittal posture of the forelimbs (Fig. 5). The mobile and pivotal clavicle evolved before the divergence of multituberculates and therians. The parasagittal forelimb posture of living therians was not present in such

archaic therians as *Zhangheotherium*, *Henkelotherium*¹⁶ and *Vincelestes*¹⁷, and it is best considered to have arisen later in therian evolution.

Although the cochleae of living monotremes and therian mammals all have some degree of coiling, it remains unclear whether a coiled cochlea evolved only once or in parallel among monotremes and therians. In early marsupials and placentals^{33,36}, the promontorium that houses the cochlea is inflated and bulbous as a result of the cochlear coiling. Within the cochlear canal, the osseous spiral laminae of the bony labyrinth are well developed to support the membranous labyrinth³⁶. In contrast, the cochleae of monotremes lack such internal bony structures^{28,33}.

Because the finger-like promontorium is closely correlated to an uncoiled cochlea in diverse early mammals^{30,31,33–35}, its presence in *Zhangheotherium* suggests that the coiled cochlea and the inflated promontorium of more derived therians^{17,33,36} are not present in symmetrodonts. Within the framework of mammalian phylogeny supported by many independent characters (Fig. 5), the coiled cochlea of therians would best be considered to have evolved later in such derived mammals as *Vincelestes*¹⁷, marsupials and placentals, independently of the partially coiled cochlea of monotremes (Fig. 5).

Conclusions

The nearly complete skeleton of *Zhangheotherium quinquecuspidens* has yielded new and more comprehensive anatomical information about the early therian mammals. New evidence on basicranial and postcranial anatomy from *Zhangheotherium* corroborates the hypothesis that symmetrodonts are a part of the basal therian radiation. Postcranial features of this new therian mammal support a sister-group relationship between multituberculates and therian mammals. A mobile clavicle–interclavicle joint that allows a wide range of movement of the forelimb has an ancient origin in the mammalian phylogeny. The abducted forelimb inferred for *Zhangheotherium* and other archaic therians suggests that early therian mammals lacked the more parasagittal forelimb posture of most living therians. The presence of a finger-like promontorium in *Zhangheotherium* indicates, albeit indirectly, that an uncoiled cochlea was present in symmetrodonts and that the coiled cochlea was a development later in therian evolution. □

Methods

Phylogeny of mammals. Based on a most parsimonious tree from PAUP (3.1.1. Exhaustive search) from 66 dental, cranial and postcranial characters that can be examined in *Zhangheotherium* (see Supplementary information for the character list and matrix). The most parsimonious tree has: tree length = 108, consistency index = 0.759, retention index = 0.840. Numbers on the branches represent the bootstrap value of 100 bootstrap replicas (the 50% majority bootstrap consensus tree and the most parsimonious tree have identical topology).

Abbreviated apomorphy list. Node A (Mammalia of ref. 43 or Mammalia-formes of ref. 27): prezygapophysis absent on axis; shallow patellar groove present on femur; notches for quadrate and quadratojugal absent in squamosal; promontorium present, unilateral occlusion of lower jaw; rotation of lower jaw during occlusion; differentiation of postcanine crowns into premolars and molars. Equivocal apomorphies: proatlantal neural arch absent in adults; number of divided postcanine roots is not more than three. Node B (*Gobiconodon* plus the crown group of mammals): atlas ribs fused in adults; postaxial ribs fused in adults; sesamoid bone present in the pedal flexor tendon; a distinctive mandibular foramen fully formed; meckelian groove is vestigial or absent in adult; glenoid of scapula concave (instead of saddle shaped) and facing posteroventrally. Equivocal apomorphies: acetabular dorsal emargination closed; presence of external pedal (tarsal) spur; foramen for superior ramus of stapedial artery enclosed. Node C (the crown group of mammals, or the Mammalia of ref. 27): presence of well-developed patellar groove on femur; greater rotation of the mandible during occlusion; main cusps of the molars arranged in reversed triangle; more transverse orientation of protocristid. Equivocal apomorphies: absence of meckelian groove on the mandible

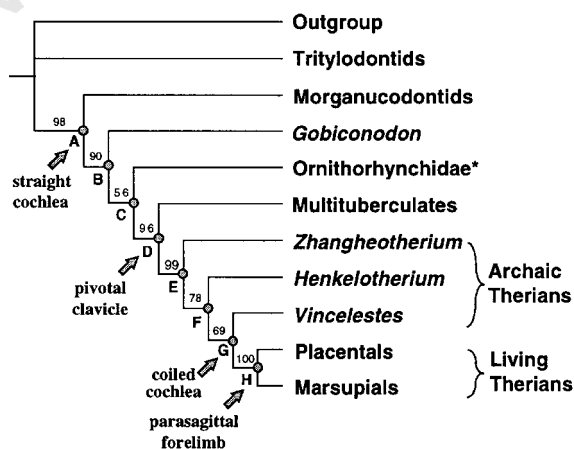


Figure 5 Phylogenetic relationships of *Zhangheotherium quinquecuspidens*. Asterisk: given this phylogeny, the coiled membranous labyrinth without corresponding coiling of the bony labyrinth in monotremes^{28,33} is considered to be a convergence to the coiled cochlea of derived therians. See Methods for details of nodes A–G.

(reversed in *Zhangheotherium* and *Henkelotherium*); main cusps of molar arranged in reverse triangle (lost in multituberculates). **Node D** (Theriiformes of ref. 27): mobile joint between clavicle and sterno-interclavicular apparatus; supraspinous fossa present on scapula; strong acromion extending below the level of glenoid of scapula; coracoid reduced to a process and fused to scapula; glenoid of scapula narrower than humeral head and facing posteroventrally; humeral and femoral heads spherical and strongly inflected; entepicondyle and ectepicondyle of humerus more reduced; styloid process of radius better developed; greater trochanter of femur directed dorsally; calcaneal tubercle elongate; presence of a separate peroneal tubercle; incipient superposition of astragalus over calcaneum; sesamoid bones in pedal flexor tendon paired; epitympanic recess in petrosal fully developed. Equivocal apomorphies: anterior sternal element (procoracoid) either fused with the manubrium sterni or lost; lesser trochanter of femur oriented ventromedially or ventrally; distinct tibial malleolus and fibular styloid process. **Node E** ('holotheres' of ref. 43): supraspinous fossa developed to the full length of scapula; flat medial surface of scapula; greater tubercle of humerus larger than the lesser; intertubercular groove very deep and narrow; squamosal with a postglenoid depression; cranial moiety of squamosal broad; post-tympanic recess present; craniomandibular joint anterior to fenestra vestibuli; crista interfenestralis limited to promontorium; squamosal wall flanking the epitympanic recess. Equivocal apomorphies: distal humerus forming a trochlea for ulnar articulation. **Node F**: talonid present on lower molar; wear facet present on talonid; protocristid of lower molar very transverse. Equivocal apomorphies: interclavicle absent in adults. **Node G** (Tribosphenida of ref. 43): upper molar with protocone; coronoid absent in adult; meckelian groove absent in adults. **Node H** (crown group of marsupials and placentals, or the Theria of ref. 27): manubria sterni small; the torsion of humerus weak; ulnar condyle on the humerus lost; fibula lost contact to calcaneum; cochlea elongate and coiled at least 360 degrees; lower molar talonid fully developed with a basin. Equivocal apomorphies: fusion of basilar neural arch. **Gobiconodon**: equivocal apomorphy: ulnar articulation on distal humerus forming a trochlea (convergent to therians). **Ornithorhynchidae**: equivocal apomorphies: postcanines with multiple roots (convergent to tritylodontids). Autapomorphy: lower molars with anteroposteriorly compressed talonid. **Multituberculates**: equivocal apomorphy: meckelian groove absent in adults. Autapomorphies: deep peroneal groove.

Received 16 July; accepted 22 September 1997.

- Wang, Y.-Q., Hu, Y.-M., Zhou, M.-Z. & Li, C.-K. in *Sixth Symposium on Mesozoic Terrestrial Ecosystems and Biota, Short Papers* (eds Sun, A.-L. & Wang, Y.-Q.) 221–227 (China Ocean Press, Beijing, 1995).
- Chen, P.-J. et al. Studies on the Late Mesozoic continental formations of western Liaoning. *Bull. Nanjing Inst. Geol. Palaeontol. Acad. Sin.* **1**, 22–55 (1980). (In Chinese.)
- Chen, P.-J. & Chang, Z.-L. Nonmarine Cretaceous stratigraphy in eastern China. *Cretaceous Res.* **15**, 245–257 (1994).
- Jin, F. New advances in the Late Mesozoic stratigraphic research of Western Liaoning, China. *Vert. Palaeontol.* **34**, 102–122 (1996).
- Hou, L.-H., Zhou, Z.-H., Martin, L. D. & Feduccia, A. A beaked bird from the Jurassic of China. *Nature* **377**, 616–618 (1995).
- Hou, L.-H., Martin, L. D., Zhou, Z.-H. & Feduccia, A. Early adaptive radiation of birds: evidence from fossils from Northeastern China. *Science* **274**, 1164–1165 (1996).
- Ji, Q. & Ji, S.-A. *Protarchaeopteryx*, a new genus of Archaeopterygidae of China. *Chin. Geol.* **1997** (3), 38–41 (1997). (In Chinese.)
- Ji, Q. & Ji, S.-A. Discovery of the earliest bird fossils in China and the origin of birds. *Chin. Geol.* **1996** (10), 30–33 (1996). (In Chinese.)
- Smith, P. E. et al. Date and rates in ancient lakes: ^{40}Ar – ^{39}Ar Evidence for an early Cretaceous age for the Jehol group, Northeast China. *Can. J. Earth Sci.* **32**, 1426–1431 (1995).
- Simpson, G. G. *A Catalogue of the Mesozoic Mammalia in the Geological Department of the British Museum* (Oxford Univ. Press, London, 1928).
- Clemens, W. A. Late Jurassic mammalian fossils in the Sedgwick Museum, Cambridge. *Palaeontology* **6**, 373–377 (1963).
- Cassiliano, M. L. & Clemens, W. A. Jr in *Mesozoic Mammals: the First Two-thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 150–161 (Univ. California Press, Berkeley, 1979).
- Jenkins, F. A. Jr & Parrington, F. R. Postcranial skeleton of the Triassic mammals *Eozostrodon*, *Megazostrodon*, and *Erythrotherium*. *Phil. Trans. R. Soc. Lond. B* **273**, 387–431 (1976).
- Krause, D. W. & Jenkins, F. A. Jr The postcranial skeleton of North American multituberculates. *Bull. Mus. Comp. Zool.* **150**, 199–246 (1983).
- Jenkins, F. A. Jr & Schaff, C. R. The Early Cretaceous mammal *Gobiconodon* (Mammalia Triconodonts) from the Cloverly Formation in Montana. *J. Vert. Paleontol.* **6**, 1–24 (1988).

- Krebs, B. Das Skelett von *Henkelotherium guimarotae* gen. et sp. nov. (Eupantotheria Mammalia) aus dem Oberen Jura von Portugal. *Berlin. Geowiss. Abh. A* **133**, 1–110 (1991).
- Rougier, G. W. *Vincelestes neuquenianus* Bonaparte (Mammalia, Theria), un primitivo mammifero del Cretacico Inferior de la Cuenca Neuquina. Thesis, Univ. Nacional de Buenos Aires (1993).
- Sereni, P. & McKenna, M. C. Cretaceous multituberculate skeleton and the early evolution of the mammalian shoulder girdle. *Nature* **377**, 144–147 (1995).
- Fox, R. C. Upper molar structure in the Late Cretaceous symmetrodont *Symmetrodontoides* Fox, and a classification of the Symmetrodonta (Mammalia). *J. Paleontol.* **59**, 21–26 (1985).
- Cifelli, R. Cretaceous mammals of southern Utah. III. Therian mammals from the Turonian (Early Late Cretaceous). *J. Vert. Paleontol.* **10**, 332–345 (1990).
- Kermack, K. A., Mussett, F. & Rigney, H. W. The skull of *Morganucodon*. *Zool. J. Linn. Soc.* **71**, 1–158 (1981).
- Crompton, A. W. & Luo, Z. in *Mammal Phylogeny* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 30–44 (Springer, New York, 1993).
- Rougier, G. W., Wible, J. R. & Hopson, J. A. Reconstruction of the cranial vessels in the Early Cretaceous mammal *Vincelestes neuquenianus*: implications for the evolution of the mammalian cranial vascular system. *J. Vert. Paleontol.* **12**, 188–216 (1992).
- Marshall, L. G. & Muizon, C. de in *Pucadelphys andinus* (Marsupialia, Mammalia) from the early Paleocene of Bolivia (ed. Muizon, C. de) *Mém. Mus. Natl. Hist. Nat. Paris* **165**, 21–90 (1995).
- Luo, Z. & Crompton, A. W. Transformations of the quadrate (incus) through the transition from non-mammalian cynodonts to mammals. *J. Vert. Paleontol.* **14**, 341–374 (1994).
- Rougier, G. W., Wible, J. R. & Hopson, J. A. Basicranial anatomy of *Priacodon fruitaensis* (Triconodontidae, Mammalia) from the Late Jurassic of Colorado, and a reappraisal of mammalia-form interrelationships. *Am. Mus. Novit.* **3183**, 1–28 (1996).
- Rowe, T. 1988. Definition, diagnosis, and origin of Mammalia. *J. Vert. Paleontol.* **8**, 241–264 (1988).
- Zeller, U. Die Entwicklung und Morphologie des Schädels von *Ornithorhynchus anatinus* (Mammalia: Prototheria: Monotremata). *Abh. Senckenb. Naturf. Gesell.* **545**, 1–188 (1989).
- Wible, J. R. Origin of Mammalia: the craniodental evidence reexamined. *J. Vert. Paleontol.* **11**, 1–28 (1991).
- Luo, Z., Crompton, A. W. & Lucas, S. G. Evolutionary origins of the mammalian promontorium and cochlea. *J. Vert. Paleontol.* **15**, 113–121 (1995).
- Graybeal, A., Rosowski, J., Ketten, D. R. & Crompton, A. W. Inner ear structure in *Morganucodon*, an early Jurassic mammal. *Zool. J. Linn. Soc.* **96**, 107–117 (1989).
- Kielan-Jaworowska, Z., Presley, R. & Poplin, C. The cranial vascular system in taeniolabidoid multituberculate mammals. *Phil. Trans. R. Soc. Lond. B* **313**, 525–602 (1986).
- Luo, Z. & Ketten, D. R. CT scanning and computerized reconstructions of the inner ear structure of multituberculate mammals. *J. Vert. Paleontol.* **11**, 220–228 (1991).
- Meng, J. & Wyss, A. Monotreme affinities and low-frequency hearing suggested by multituberculate ear. *Nature* **377**, 141–144 (1995).
- Lillegraven, J. A. & Krusat, G. Cranio-mandibular anatomy of *Haldanodon expectatus* (Docondontia; Mammalia) from the Late Jurassic of Portugal and its implications to the evolution of mammalian characters. *Contrib. Geol.* **28**, 39–138 (1991).
- Meng, J. & Fox, R. C. Therian petrosals from the Oldman and Milk River Formations (Late Cretaceous), Alberta, Canada. *J. Vert. Paleontol.* **15**, 122–130 (1995).
- Wible, J. R. & Hopson, J. A. in *Mammal Phylogeny* Vol. 1 (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 45–62 (Springer, New York, 1993).
- Kielan-Jaworowska, Z. & Gambaryan, P. P. Postcranial anatomy and habits of Asian multituberculate mammals. *Fossils & Strata* **36**, 1–92 (1994).
- Jenkins, F. A. Jr The movement of the shoulder in clavicate and a clavicate mammals. *J. Morphol.* **144**, 71–84 (1974).
- Jenkins, F. A. Jr & Weis, W. A. The functional anatomy of the shoulder in the Virginia opossum (*Didelphis virginiana*). *J. Zool.* **188**, 379–410 (1979).
- Jenkins, F. A. Jr The functional anatomy and evolution of the mammalian humero-ulnar joint. *Am. J. Anat.* **137**, 281–298 (1973).
- Gambaryan, P. P. & Kielan-Jaworowska, Z. Sprawling versus parasagittal stance in multituberculate mammals. *Acta Palaeontol. Pol.* **42**, 13–44 (1997).
- Hopson, J. A. in *Major Features of Vertebrate Evolution* (eds Prothero, D. R. & Schoch, R. M.) 190–219 (Paleontological Society Short Courses, Knoxville, TN, 1994).
- Kielan-Jaworowska, Z., Crompton, A. W. & Jenkins, F. A. Jr The origin of egg-laying mammals. *Nature* **326**, 871–873 (1987).
- Crompton, A. W. in *Early Mammals* (eds Kermack, D. M. & Kermack, K. A.) 65–87 (Academic, London, 1971).
- Crompton, A. W. & Jenkins, F. A. Jr in *Mesozoic Mammals: the First Two-thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 59–72 (Univ. California Press, Berkeley, 1979).
- Kemp, T. S. The relationships of mammals. *Zool. J. Linn. Soc.* **77**, 353–384 (1983).
- Rowe, T. in *Mammal Phylogeny* Vol. 1 (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 129–145 (Springer, New York, 1993).
- Klima, M. Die Frühentwicklung des Schultergürtels und des Brustbeins bei den Monotremen (Mammalia: Prototheria). *Adv. Anat. Embryol. Cell Biol.* **47**, 1–80 (1973).
- Klima, M. Early development of the shoulder girdle and sternum in marsupials (Mammalia: Metatheria). *Adv. Anat. Embryol. Cell Biol.* **109**, 1–91 (1987).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank K. C. Beard, M. Dawson, R. Fox, Z. Kielan-Jaworowska, B. Krebs, T. Martin, J. Meng, R. Presley, G. Rougier, P. Sereno, A. Sun and J. Wible for suggestions on the manuscript; L.-H. Hou for field assistance; G.-H. Cui for photography; and H. Zhang for preparation. Research was supported by National Natural Science Foundation of China (to C.L.), the Rea Postdoctoral Fellowship of Carnegie Museum (to Y.W.), National Science Foundation of USA, National Geographic Society, and the M. Graham Netting Fund of Carnegie Museum (to Z.L.).

Correspondence and requests for materials should be addressed to: Z. Luo (e-mail: luoz@clpgh.org) or C. Li (e-mail: lihwpp@midwest.com.cn).

Photonic-bandgap microcavities in optical waveguides

J. S. Foresi*, P. R. Villeneuve†, J. Ferrera‡, E. R. Thoen‡, G. Steinmeyer‡, S. Fan†, J. D. Joannopoulos†, L. C. Kimerling*, Henry I. Smith‡ & E. P. Ippen‡†

* Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

‡ Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Confinement of light to small volumes has important implications for optical emission properties: it changes the probability of spontaneous emission from atoms, allowing both enhancement and inhibition. In photonic-bandgap (PBG) materials^{1–4} (also known as photonic crystals), light can be confined within a volume of the order of $(\lambda/2n)^3$, where λ is the emission wavelength and n the refractive index of the material, by scattering from a periodic array of scattering centres. Until recently^{5,6}, the properties of two- and three-dimensional PBG structures have been measured only at microwave frequencies. Because the optical bandgap scales with the period of the scattering centres, feature sizes of around 100 nm are needed for manipulation of light at the infrared wavelength (1.54 μm) used for optical communications. Fabricating features this small requires the use of electron-beam or X-ray lithography. Here we report measurements of microcavity resonances in PBG structures integrated directly into a sub-micrometre-scale silicon waveguide. The microcavity has a resonance at a wavelength of 1.56 μm , a quality factor of 265 and a modal volume of 0.055 μm^3 . This level of integration might lead to new photonic chip architectures and devices, such as zero-threshold microlasers, filters and signal routers.

A single mode waveguide is used to both confine light in the PBG structure and control the polarization of the light. A diagram of the waveguide structure is shown in Fig. 1a. The waveguide consists of a silicon (Si) strip on a silicon dioxide (SiO_2) layer. The low-index SiO_2 layer separates the guided mode from the underlying Si substrate and allows strong field confinement in the Si waveguide. The introduction of a periodic array of holes into the waveguide limits the wavevector to π/a , where a is the spatial period, or lattice constant, of the array of holes. This allowable range of wavevectors is similar to the Brillouin zone used in solid-state physics. As well as putting a limit on the wavevector, the periodic array of holes (that is, the PBG structure) has the effect of folding the dispersion relation of the strip waveguide and of splitting the lowest-order mode to form two allowable guided modes⁷. The splitting at the Brillouin zone edge is called a bandgap; its size is determined by the dielectric contrast of the periodic structure. For the device shown in Fig. 1a, the dielectric contrast is defined by the Si waveguide ($\epsilon_{\text{Si}} = 12.1$) and the air holes ($\epsilon_{\text{air}} = 1$) etched into the waveguide. The large contrast opens a gap that is 27% of the midgap frequency. With a midgap wavelength of $\lambda = 1.54 \mu\text{m}$, the gap is 400 nm in width. This gap is significantly larger than that which can be achieved with surface-corrugated waveguides.

If a defect is included in the PBG structure, a state can be formed in the gap. This state is analogous to a defect or impurity state in a semiconductor that forms a level within the semiconductor bandgap. Defects in PBG structures are engineered by incorporating breaks in the periodicity of the PBG device. A break in periodicity leads to a defect state that is localized in real space and therefore extended in wavevector space. Controlling the coupling of this

extended state to radiation modes is necessary to optimize the performance of the defect as a resonator or filter. Taking these considerations into account, we designed a PBG microcavity for operation near $\lambda = 1.54 \mu\text{m}$. The dimensions of the device, and the computed transmission through the device, are shown in Fig. 1a, b. The computation shows a wide stop-band from 1,300 to 1,700 nm. The resonance is predicted to have a wavelength of 1,547 nm and a cavity quality factor Q of 280. The Q of the cavity is defined as $\lambda/\Delta\lambda$ of the cavity resonance and is a measure of the ratio of the optical energy stored in the microcavity to the cycle-average power radiated out of the cavity. The resonant mode is strongly confined within the microcavity. The modal volume, V , is defined as

$$VP_{\text{max}} = \int P(r) d^3r \quad (1)$$

where $P(r)$ is the total electromagnetic energy density of the mode, and P_{max} is the peak value of $P(r)$. With this definition, a modal volume of 0.055 μm^3 is computed.

The minimum feature size associated with the waveguide microcavity occurs between the edges of the holes and the edge of the waveguide. For the device shown in Fig. 1a, the minimum feature size is 0.13 μm , which is currently beyond the limits of optical lithography. To achieve pattern transfer at this length scale, we used X-ray lithography with a Cu_L source at 1.3 nm; the mask was made by electron-beam lithography⁸. This approach easily resolved the 0.13 μm PBG feature. A scanning electron micrograph of a processed device is shown in Fig. 2. This device was fabricated from a Unibond silicon-on-insulator substrate with a 0.2 μm single-crystal Si layer on a 1.0 μm SiO_2 layer.

For testing purposes, we designed the waveguide to be curved at both the input and output ports. This curvature allows the input and output of the waveguide to be offset during the testing. Any light that couples through the substrate is spatially separated from the waveguide mode. Flares at both the input and output widen the guide to 5.0 μm to improve coupling into the waveguide. Total waveguide lengths of 1.5 mm were used.

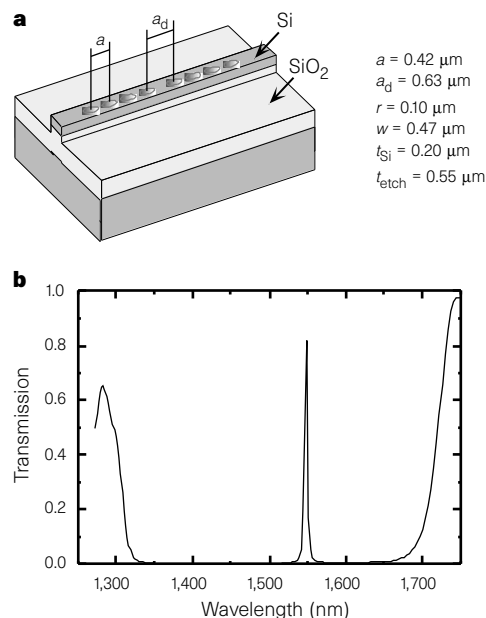


Figure 1 Dimensions and basic properties of a PBG waveguide microcavity. **a**, Schematic of a PBG waveguide microcavity with dimensions for operation at $\lambda = 1.54 \mu\text{m}$: a is the hole period, a_d is the defect length, r is the hole radius, w is the waveguide width, t_{Si} is the silicon thickness and t_{etch} is the total etch depth through both the Si and SiO_2 . **b**, Computation of the transmission through the structure. The transmission spectrum corresponds to the device dimensions listed. The resonance is centred at 1,547 nm and has a Q of 280.

To test the devices, a continuous-wave colour-centre laser ($\text{NaCl}:\text{OH}^-$) with a tuneable range from 1,500 to 1,625 nm is used. This laser has ~ 250 mW of output power across the tuning range. An optical multichannel analyser provides continuous monitoring of the wavelength during the testing, and a portion of the beam is split off to monitor the laser power. The laser beam is focused into an optical fibre that has a lensed output tip positioned to couple light into the waveguide. A $\times 40$ objective lens collects the transmitted light at the waveguide output. A polarizing beam splitter allows discrimination between TE and TM polarized light. The $\times 40$ objective images the waveguide output through the beam splitter onto a $100\text{ }\mu\text{m}$ pinhole. This pinhole reduces the amount of substrate-coupled light that reaches the detector. The wavelength reference, power monitor and transmitted power are all collected simultaneously by a data-acquisition system.

The measured transmission spectrum for a PBG waveguide microcavity with four holes on each side of the defect is shown as the solid line in Fig. 3; the theoretical transmission for this particular device is shown as a dotted line. At resonance, a sharp peak in transmission is seen. The measured Q of this resonance is 265, compared with the calculated Q of 280. This resonance wavelength is 1,560 nm, a 1% shift from the calculated resonance of 1,547 nm. The measured values are in remarkable agreement with

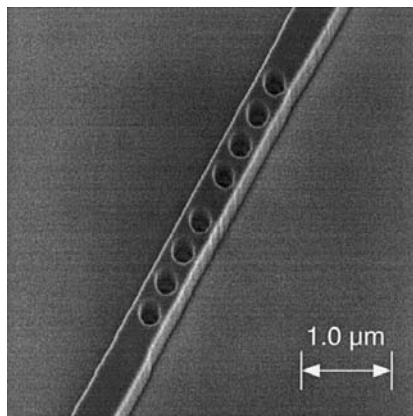


Figure 2 Scanning electron micrograph of a PBG waveguide microcavity fabricated by X-ray lithography. The pattern was transferred to the Unibond silicon-on-insulator substrate with a Cr lift-off mask. The Si etching was performed in a plasma of CF_4 and O_2 ; the SiO_2 etching used a CHF_3 plasma. After etching, the Cr was removed and the samples were cut and polished for testing.

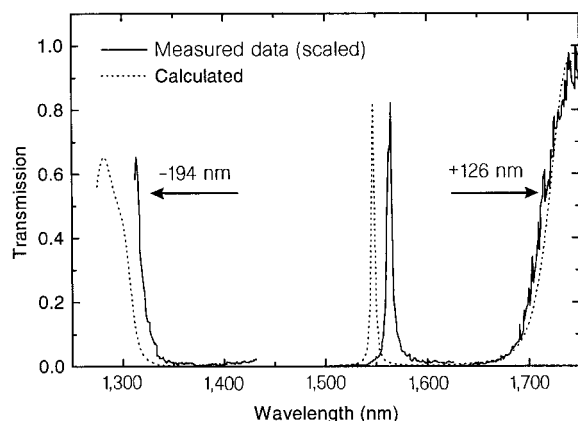


Figure 3 Comparison of measured transmission (solid lines) to calculated transmission (dotted line) for a PBG microcavity with four holes on each side of the microcavity. The upper-band and lower-band edge transmission data have been shifted in wavelength as described in the text.

theory. The transmission spectrum was calculated from dimensions measured on the actual device. The discrepancy between calculated and measured transmission features is well within the accuracy of the scanning electron microscopy technique used to determine the device dimensions. As mentioned earlier, the modal volume is $0.055\text{ }\mu\text{m}^3$. This is the most spatially localized resonant photon state ever achieved and corresponds to a volume of only five times $(\lambda/2n)^3$.

Because the tuning range of the colour-centre laser is smaller than the PBG stop-band, not all of the PBG features (resonance and band edges) could be measured on a single waveguide microcavity. However, the scalability of the PBG properties with wavelength allows tuning of the device dimensions to move the desired transmission features into the laser tuning range. We used this property to design devices with band edges within the tuning range. To see the high-frequency band edge we increased all of the dimensions, and to see the low-frequency band edge we decreased all of the dimensions. The scaling essentially shifts the transmission spectrum of Fig. 1b to longer or shorter wavelengths, but the overall shape of the spectrum does not change. It should be noted that, because all of the microcavities are processed on the same substrate, the Si thickness of the band-edge devices could not be tuned with the other dimensions.

To map all of the data onto a single plot, we performed transmission simulations using the measured dimensions of the band-edge device. The theoretical shifts were calculated and applied to the measured data. The computations take into account the deviation of the device thickness from perfect scaling, and show that the thickness variation results only in a modification of the wavelength shift, not a change in the shape of the transmission characteristic. For the upper-frequency band edge (that is, short wavelength) a shift of -194 nm was determined, and for the lower-frequency band edge a shift of $+126$ nm was found. The scaled data are shown in Fig. 3. Excellent agreement between theory and experiment is seen. The upper band edge is red-shifted by 1%, and the lower band edge overlaps the prediction almost exactly. Simulations indicate that the low-frequency band edge does not shift appreciably with dimensional variations, whereas the resonance and high-frequency band edges are more sensitive to changes in dimension. The transmission measurements marked the first complete evaluation of PBG guided mode properties, including resonant states, at optical frequencies.

The resonance wavelength of the PBG waveguide microcavity can be altered by changing the size of the defect³. The results shown in Figs 1–3 were obtained with a defect length a_d of $1.50a$, where a is the period of the PBG holes. In Fig. 4 we show measured resonances

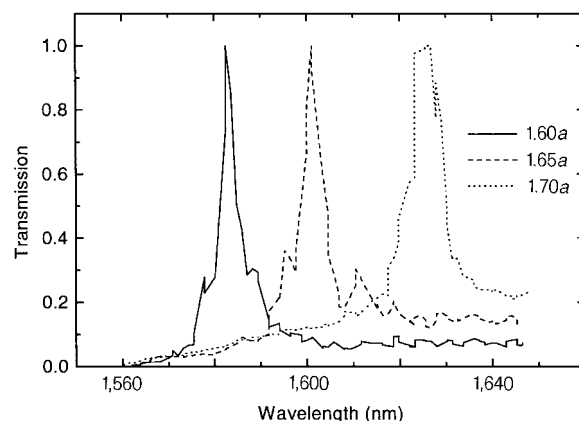


Figure 4 Transmission characteristic of PBG waveguide microcavities for various defect lengths. Three different defect lengths are shown. The resonance peak shifts to longer wavelengths with increasing defect length.

for defects ranging in length from $1.60a$ to $1.70a$. Two phenomena are evident. As the defect length increases, the resonance wavelength shifts to longer values, and the cavity Q is reduced. The increase in cavity length results in a corresponding increase in the effective index 'seen' by the resonant mode, reducing the frequency of the mode. This reduction enhances the coupling of the resonant mode to the waveguide mode, which, in turn, increases the cycle-average radiation out of the cavity and hence reduces the Q .

To study the PBG microcavities, we concurrently developed submicrometre optical waveguides⁹. The reduction in size achieved with high-dielectric contrast materials (these waveguides have cross-sectional dimensions more than an order of magnitude smaller than common low-index-difference waveguides) can lead to densely integrated waveguide structures for optical communication systems. This waveguide technology could lead to an integrated optics approach compatible in materials and dimensions with integrated electronics.

The experimental verification of PBG theory presented above provides confidence that this technology is realizable at optical communication frequencies and can pave the way to the development of novel devices. Because of its high Q/V ratio, the PBG waveguide microcavity can be considered for use in spontaneous-emission-enhancement devices, such as resonant-cavity light-emitting diodes. The control of spontaneous emission is of interest for increasing the light output and narrowing the linewidth of light-emitting diode structures, and for reducing the lasing threshold of semiconductor lasers. By coupling an optical transition to the microcavity resonance, the spontaneous emission rate can be enhanced by a factor η over the rate with no cavity. The expression for η is given as¹⁰

$$\eta = \frac{Q}{4\pi V} \left(\frac{c}{n\nu} \right)^3 \quad (2)$$

where c is the speed of light and ν is the optical transition frequency. For the specific PBG waveguide microcavity shown in Fig. 2, this enhancement factor is calculated to be 35, which is significantly larger than any enhancement yet measured. This large spontaneous-emission enhancement could lead to the development of zero-threshold laser devices and efficient erbium-doped Si light emitters. An additional advantage of the PBG waveguide microcavity over more conventional stacked mirror microcavities is the large band-gap. With a 400 nm stop-band, it is possible to have only a single resonant mode in the gain bandwidth of a semiconductor device. The single-mode resonance reduces the available pathways for optical emission away from the resonant wavelength, and channels all of the available optical power into the resonant mode. In contrast with vertically integrated resonant-cavity devices, the configuration used here allows a light-emitting cavity to be directly coupled into a waveguide, so that light collection from the cavity is highly efficient. □

Received 23 July; accepted 25 September 1997.

- Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. Photonic crystals: putting a new twist on light. *Nature* **386**, 143–149 (1997).
- Soukoulis, C. M. (ed.) *Photonic Band Gap Materials* (Kluwer, Dordrecht, 1996).
- Joannopoulos, J. D., Meade, R. D. & Winn, J. N. *Photonic Crystals* (Princeton, New York, 1995).
- Yablonovitch, E. Photonic band-gap structures. *J. Opt. Soc. Am. B* **10**, 283–295 (1993).
- Kraus, T., De La Rue, R. & Band, S. Two-dimensional photonic bandgap structures operating at near-infrared wavelengths. *Nature* **383**, 699–702 (1996).
- Grünung, U., Lehmann, V., Ottow, S. & Busch, K. Macroporous silicon with a complete two-dimensional photonic bandgap centered at 5 μm . *Appl. Phys. Lett.* **68**, 747–749 (1996).
- Fan, S., Winn, J. N., Devenyi, A., Chen, J. C., Meade, R. D. & Joannopoulos, J. D. Guided and defect modes in periodic dielectric waveguides. *J. Opt. Soc. Am. B* **12**, 1267–1272 (1995).
- Yang, I. Y. et al. Combining and matching optical, e-beam and x-ray lithographies in the fabrication of Si CMOS circuits with 0.1 and sub-0.1 μm features. *J. Vacuum Sci. Technol. B* **13**, 2741–2744 (1995).
- Foresi, J. S. thesis, MIT (1997).
- Yokoyama, H. & Brorson, S. D. Rate equation analysis of microcavity lasers. *J. Appl. Phys.* **66**, 4801–4805 (1989).

Acknowledgements. This work was supported by the Air Force Office of Scientific Research (AFOSR), the National Science Foundation (NSF-MRSEC) and the Army Research Office (ARO).

Correspondence and requests for materials should be addressed to P.R.V. (e-mail: pville@mit.edu).

Branched fractal patterns in non-equilibrium electrochemical deposition from oscillatory nucleation and growth

Vincent Fleury

Laboratoire de Physique de la Matière Condensée, École Polytechnique, 91128 Palaiseau cedex, France

The branched fractal structures formed by non-equilibrium electrodeposition of metals¹ have for several years been considered as model systems for the study of branching and fractal growth processes generally^{2–6}. Most studies have focused on the large-scale structure of the deposits, but the question of how the branching pattern emerges from the nucleation and growth of the polycrystalline metal at the microscopic scale remains unclear. Here I present experimental and theoretical results which suggest that branched electrodeposits may arise from an oscillatory character in the nucleation kinetics. For this kind of deposition, nucleation is probabilistic but biased towards higher electric fields. I suggest that a given nucleation event is followed by a recovery phase before a subsequent event is possible. This oscillatory nature generates a polycrystalline deposit, the grain size of which determines the level of 'noise' which is amplified by the familiar laplacian ('fingering') instabilities of non-equilibrium growth^{2–6} into a macroscopically fractal structure.

Fractal models of growth^{2–4} assume that single 'grains' of the same size diffuse and aggregate (the so-called diffusion-limited aggregation model⁶) or nucleate probabilistically on the border of the structure (the dielectric breakdown model⁷). However, true polycrystalline fractal structures are made of grains of a given size, which neither diffuse nor have their final radius instantly at nucleation. I now investigate how the grain texture at the small scale of the deposit is connected to the branched structure of the large scale, and to the fields driving the growth. In particular, how does the growth speed of individual grains match the average growth speed of the branched structure at the larger scale? What determines the size of the grains and the formation of new nucleation events?

The experimental method consists of growing branched metallic structures by electrochemical deposition (ECD), and studying them from the scale of individual grains. For this I have used a new type of electrochemical cell. Glass slides are covered with a thin gold coating and used as support for the growth. Two gold edges, 1,000 Å thick, are deposited 2 cm apart on the support. Copper foil is used as spacers on the cathodic and anodic sides. The deposit starts growing from the 1,000 Å gold cathode, and keeps growing on the slide. The deposit remains on the substrate after cell rinsing and can be used for further analyses. The growth was performed in cells of thicknesses 15 μm , 50 μm and 100 μm . Solutions of CuSO_4 of concentrations 0.01, 0.02, 0.04 and 0.1 were used. Constant current was applied in the range 1–100 μA . With this set-up, I obtained robust branched deposits which could be observed *in situ* up to a magnification of $\times 400$, and *ex situ* in the scanning electron microscope (SEM) and atomic force microscope (AFM). The growth was recorded on VHS tape with a CCD (charge-coupled device) camera, and analysed with NIH Image software.

The measured growth speed of the deposit in the range of parameters studied is roughly equal to that predicted by Chazalviel's theory¹⁸ (Table 1), which is the recession speed of the anions^{8–11} in the bulk. The SEM (Fig. 1) shows that the branches are made of a

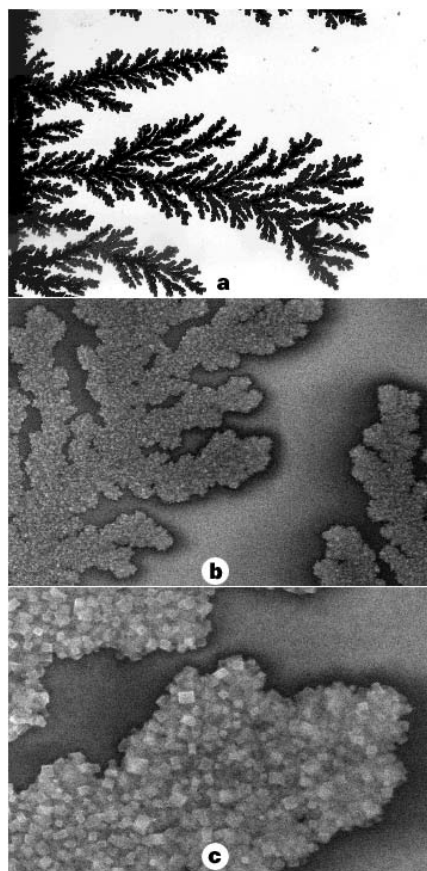


Figure 1 A typical deposit in a cell 100 μm thick, with copper sulphate 0.01 M, current 5 μA . **a**, Magnification $\times 45$; **b**, $\times 1,780$; **c**, $\times 7,580$. As one zooms in on the structure, the fine polycrystalline features appear. The branched DLA-like deposit is made of a carpet of grains which are sitting side by side. Of course, single grains do not diffuse or migrate in the solution. Copper ions arrive close to the deposit by migration and diffusion (which, in this régime, are of comparable magnitudes) and deposit on individual grains which nucleate and grow repeatedly on previous grains. There is a characteristic roughness, given by the size of the grains.

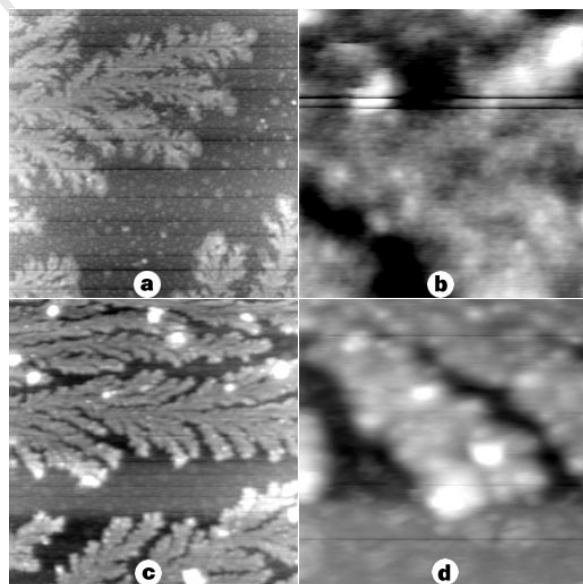


Table 1 Values of the growth speed for different conditions

Current (μA)	15 μm cell		50 μm cell		
	Exp. ($\mu\text{m s}^{-1}$)	Theor. ($\mu\text{m s}^{-1}$)	Current (μA)	Exp. ($\mu\text{m s}^{-1}$)	Theor. ($\mu\text{m s}^{-1}$)
15	0.89	0.96	100	3.20	1.64
10	0.44	0.64	25	0.57	0.41
8	0.2	0.51	15	0.48	0.24
4	0.08	0.25	10	0.54	0.16

The comparison between the measured speeds and the prediction of the Chazalviel model⁸ is fair. Note that the value of the mobility of the anions at the different concentrations is not precisely known. I have estimated the mobility of the anions from the measurement of the electrolyte conductivity by assuming that the anionic current is 1.66 times the cationic current (which is the ratio at infinite dilution).

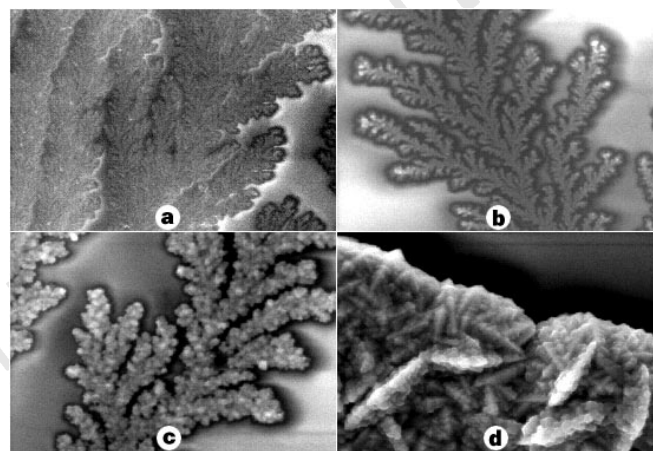


Figure 2 Texture as a function of growth speed. As the growth speed is decreased, the grain size in the deposit becomes ever larger. Cell thickness 12 μm , copper sulphate 0.02 M. **a**, 10 μA ; **b**, 8 μA ; **c**, 6 μA ; **d**, 2 μA ; all images are magnified $\times 6,380$. In **a**, the deposit is so thin and the grain is so small that it cannot be resolved, but the deposit appears to be very branched, with many small branches which clearly have the smallest cut-off. In **b**, the grain starts very faintly to appear; this is somewhat more visible towards the tips of the branches where charge effects increase the contrast. The smaller cut-off for the branch size is increased with respect to **a**. In **c**, the grain size appears very clearly, and the cut-off is again increased. In **d**, the cut-off is very coarse, and the grains are very large. In this instance, the growth was so slow that it was interrupted before the deposit had evolved into a branched structure. The as-deposited samples were imaged in the SEM without additional coatings, to avoid formation of additional grains which might have perturbed the interpretation.

Figure 3 AFM analysis of the growths at 10 μA (**a** and **b**) and 8 μA (**c** and **d**), obtained with a Park Instruments Scanning Force Microscope. The sizes of the images shown are: **a**, $6 \times 6 \mu\text{m}^2$; **b**, $0.5 \times 0.5 \mu\text{m}^2$, height range (black-white) 250 $\text{\AA}$; **c**, $6 \times 6 \mu\text{m}^2$; **d**, $1 \times 1 \mu\text{m}^2$, height range 500 $\text{\AA}$. The change in cut-off with increasing growth speed is obvious. There is also a change in grain size. For the slowest growth one can identify long whisker-like shapes several micrometres in length. For 6 μA , the typical nuclei size was assessed as about 7,000 $\text{\AA}$ (with a Dektak profilometer from Veeco technology, data not shown). However, for more rapid growth, the grain size is difficult to estimate, despite the quantitative character of the AFM. From images **b** and **d**, orders of magnitude of 100 $\text{\AA}$ and 1,000 $\text{\AA}$ are estimated for the current densities 10 μA and 8 μA (this was confirmed with the profilometer). It is worth noting that the thickness of about 20 nm which is obtained for the more space-filling deposit is in accordance with the thickness that one deduces from the assumption that all the copper present in the cell deposits on the growing irregular thin film. The AFM and SEM clearly show that the microscopic roughness is given by the grain nucleation and growth kinetics, and possible overlap. The pattern forms a fractal-like tree on the large scale. The instability does not develop as the consequence of a Mullins-Sekerka type growth of instabilities, but as probabilistic nucleation and growth of grains.

carpet of neighbouring grains stuck to the substrate. An essential observation is that the two-dimensional branches are made of three-dimensional grains. Figure 2 shows typical copper deposits grown at different speeds (that is, different currents). The grains appear larger at lower growth speed, as confirmed by AFM analysis at an even smaller scale (Fig. 3). The variation of grain size is quite important in a narrow range of current densities, and the surface roughness is on scales given by the grain size (roughly 100 Å at 10 μ A, 1,000 Å at 8 μ A and 7,000 Å at 6 μ A, as deduced from AFM analysis, surface profilometry or SEM; whiskers in the micrometre range or above were obtained at 2 μ A). The grains formed are smaller at larger growth speeds, and the typical branch size, in other words the lower cut-off for branching (and for fractal morphology), is smaller at larger growth speeds (0.1 μ m at 10 μ A, 0.5 μ m at 8 μ A, 1.5 μ m at 6 μ A; at 2 μ A, the growth was too slow to wait for the formation of actual trees, but an instability on the 10 μ m range was developing). For the slowest growth speeds, large faceted crystals are obtained¹², and the structure is not truly tree-like.

The experiment therefore shows that the growth proceeds through nucleation and growth events, with a nucleation frequency that is higher at higher currents. This nucleation process, as shown by the experiment, is random in nature, with a bias towards higher fields. The typical size of the grains, before a new nucleation event, defines the geometrical noise, and the cut-off for stability on the interface. In the range of current densities studied, a DBM-like instability exists, but the grain size is fixed by nucleation kinetics.

I propose the following theoretical model for the nucleation and growth of grains. We will only assume that a critical surface field E^* is needed for three-dimensional nucleation of a new grain. The

electrochemical equations outside the grain are well known, and can be exactly integrated⁸. The transport equations give the fluxes of cations and anions and the Poisson equation gives the electrostatic potential. Although a full three-dimensional calculation would in principle be needed, let us resort to a simple one-dimensional model: we shall grow 'rods', whose growth speed decreases as $1/R^2$ where R is the length of the rod. This simple nonlinear law aims at modelling the fact that the radial growth of the grain decreases as its surface increases. Let us remember that during electrochemical growth of deposits, a space charge forms close to the surface, owing to migration of cations and anions in opposite directions⁸. Previously, only stationary states of growth have been studied⁸. I now explain the results of the oscillatory model.

When the surface field reaches the value E^* , rapid growth of a new grain is triggered (Fig. 4); this rapid growth invades the space charge region which is found close to the surface. As a consequence, the surface field collapses, and new nucleation events are totally inhibited. While the grain grows, the growth speed decreases, until it gets much smaller than the speed of anion recession. The anions migrate away from the deposit, and the space charge progressively 'recovers'. A large field is again formed on the surface, up to the point when a new grain is 'fired', and the process repeats itself. From this model, it is seen that the grain size R_{grain} and the nucleation period T depend only on the critical field E^* and on the current j flowing through the grain: the model is mathematically closed. Figure 4a and b shows the growth for two values of E^* and the same potential across the cell: the grain size is smaller and the period is shorter for lower E^* . Figure 4a and c shows the growth for the same value of the critical field, but two values of the potential

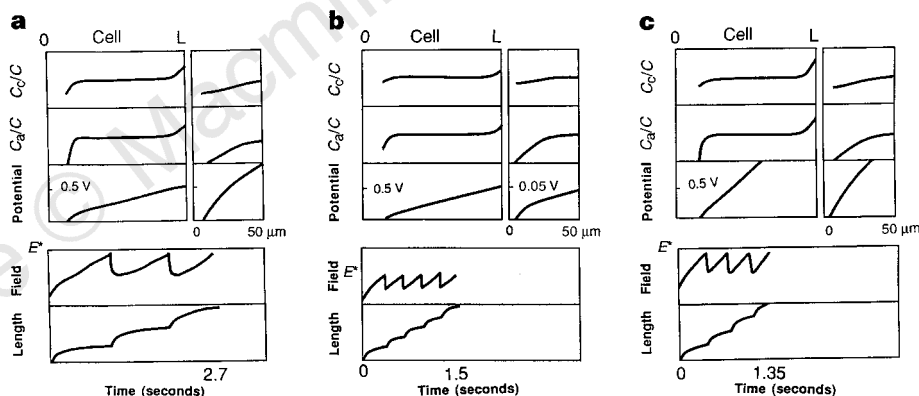


Figure 4 Growth of a line of grains deposited by nucleation and growth events. This is obtained by solving explicitly the set of transport equations for the ions, and the Poisson equation for the potential, following the work of Chazalviel⁹. The top boxes show the concentration of cations C_0 , the concentration of anions C_a and the electrostatic potential at the end of the growth across the cell (cathode to the left 0, anode to the right L). A coarse grid is used in the bulk (between 0 and L, top long boxes) and a fine grid is used close to the growing front (square boxes). The coarse grid has 100 points. The fine grid contains 50 points and it interpolates the first 5 points of the coarse mesh. The electrolyte is dilute (10^9 cm^{-3}); the total size of the coarse box is 1 mm and the resolution of the fine box is 1 μ m. These rather large scales are caused by the use of a dilute concentration. A more realistic concentration imposes exceedingly small time increments. The concentration scales of the fine and coarse boxes are the same. The potential scale of the fine box is ten times smaller than that of the coarse box (so the potential is magnified in the square boxes). Explicit recursion schemes are used for the calculation. The total simulated time in **a** is 2.7 seconds, divided into 60,000 time increments. One lets the first grain start without a field threshold. Then, when the field on the grain surface reaches E^* (4.7 V cm^{-1} in **a**), one nucleates a new grain, and lets it grow. The two bottom boxes in **a** represent the electric field facing the grains as a function of time, and the length of the deposit, obtained by integrating the flux at its tip, and taking into account a factor proportional to $1/R^2$ (R is the length of the grain). During the phase of rapid growth of a grain, the field

decreases rapidly (the positive space charge is reduced because the deposit catches up the anions). Then, as the grain growth slows down, the space charge 'recovers' because the deposit does not catch up the anions. Concomitantly, the field increases up to the threshold when a new grain nucleates. The recovery of the space charge and of the field is almost linear, because it is linked to the linear recession of the bulk anions which migrate away under the influence of the electric field. The growth speed of the line of grains is not constant but oscillates around an average. This fundamental oscillation is not linked to the growth of atomic layers, as is usual in electrochemistry (see for example ref. 18, pages 203–207), but to the recovery of the formation of the charged layer, close to a growing grain. The total length of the deposit at the end is 150 μ m. **b**, The same process as in **a** but for a smaller value of the critical field E^* (2.5 V cm^{-1}), and the same value of the potential across the cell (0.5 V). The nucleation and growth are even more conspicuous. The nucleation period and the grain size are smaller. **c**, The same critical field as in **a**, but with the potential of the anode changed to a higher value (1.5 V). In this instance, the flux is higher. The nucleation of smaller grains occurs at a faster rate. In each of **a**, **b** and **c** the concentrations and potentials are given at the end of the calculation. A numerical noise of amplitude smaller than the line thickness was smoothed out by the drawing of the interpolation. The total size of the deposit is 150 μ m. Note that, under the same current, the same distance is achieved more rapidly when the grains are smaller.

across the cell: the grain size is smaller and the period is shorter for larger fields in the bulk.

Let us now consider the entire deposit, with all the filaments growing ahead. The growth speed of the filaments is, on average, the recession speed of the anions, $\mu_a E_0$ (where μ_a is the anion mobility)⁸. This forces the growth of each line $R_{\text{grain}}/T = \mu_a E_0$. This crucial fact fixes one relationship between R_{grain} and T , independent of E^* , which is a physical constant of the surface. This implies that the current j per grain is fixed. That fixes the number of grains growing at a given moment. Once we know the surface field E^* for nucleation, the concentration of the solution, and the field E_0 across the cell, the grain size, the nucleation frequency and the number of grains, are fixed unambiguously.

The average speed is achieved by means of an oscillation of the nucleation and growth process. The period T of the oscillation is the time for recovery of the space charge, and it is also the flight time of anions away from a grain: $T = R_{\text{grain}}/\mu_a E_0$. This gives an interesting microscopic interpretation of the growth speed as found by Chazalviel⁸ using gross conservation arguments.

The surface tension could easily be incorporated into the model as an additional overpotential on the surface. Also, the transfer overpotential could be integrated into the model, but being a logarithmic correction as a function of current, it plays a secondary role. Finally, this model is consistent with existing experiments on powdery growth, especially to the extent that this proceeds via three-dimensional nucleation and growth events¹³. Also, the assumption that nucleation is triggered by a critical surface field² can be expressed in terms of the surface potential instead, and one recovers the usual probabilistic nucleation activated by an overpotential^{14–18}, with the proviso that we are dealing with a out-of-equilibrium situation.

I believe that this example gives insight into many other growth patterns, which will behave in a similar way if they form polycrystalline structures. The field that drives the long-range shape is not essential to the mechanism proposed. The essential feature is that repeated nucleation and growth of neighbouring grains behaves as a nonlinear oscillatory system with a growth speed which oscillates, at the microscopic scale, around the average speed imposed by the out-of-equilibrium constraint. The nucleation probability is biased towards higher gradients (of concentration or of potential), but the nucleation and growth kinetics are not described either by the model DLA or the DMB. The texture of the deposit at a scale where the global pattern appears discrete (grains), but space-filling (not branched), is the geometrical product of this oscillation, and it provides the geometrical noise which is further amplified by the laplacian field to give a fractal branching structure. The experiment and model are also a first step towards understanding the effect of pulsating currents on morphology of irregular deposits¹⁸. □

Received 20 November 1996; accepted 22 September 1997

1. Matsushita, M., Sano, M., Hayakawa, Y., Honjo, H. & Sawada, Y. Fractal structures of zinc metal leaves grown by electrodeposition. *Phys. Rev. Lett.* **53**, 286–289 (1984).
2. Stanley, G. & Ostrowsky, N. (eds) *On Growth and Form* NATO ASI Series Vol. 100 (Martinus Nijhoff, Kluwer Academic, Boston, 1996).
3. Viscsek, T. *Fractal Growth Phenomena*, 2nd edn (World Scientific, Singapore, 1992).
4. Meakin, P. in *Phase Transitions and Critical Phenomena* Vol. 12, (eds Domb, C. & Leibowitz, J. L.) (Academic, New York, 1988).
5. Ben-Jacob, E. & Garik, P. The formation of patterns in non-equilibrium growth. *Nature* **343**, 523–530 (1990).
6. Witten, T. A. & Sander, L. M. Diffusion limited aggregation, a kinetic critical phenomenon. *Phys. Rev. Lett.* **47**, 1400–1403 (1981).
7. Niemeyer, L., Pietronero, L. & Wiesmann, H. J. *Phys. Rev. Lett.* **52**, 1033 (1984).
8. Chazalviel, J.-N. Electrochemical aspects of the generation of ramified metallic deposits. *Phys. Rev. A* **42**, 7355–7367 (1990).
9. Fleury, V., Chazalviel, J.-N., Rosso, M. & Sapoval, B. Role of the anions in the growth speed of fractal electrodeposits. *J. Electroanal. Chem.* **290**, 249–255 (1990).
10. Fleury, V., Kaufman, J. & Hibbert, D. B. Mechanism of a morphology transition in ramified electrochemical growth. *Nature* **367**, 435–438 (1994).
11. Melrose, J. R., Hibbert, D. B. & Ball, R. C. Interfacial velocity in electrochemical deposition and the Hecker effect. *Phys. Rev. Lett.* **65**, 3009–3012 (1990).
12. Fleury, V. & Barkey, D. Runaway growth in two-dimensional electrodeposition. *Europhys. Lett.* **36**(4) 253–258 (1996).
13. Despic, A. R. & Popov, K. I. in *Modern Aspects of Electrochemistry* Vol. 7 (eds Conway, B. E. & Bockris, J. O'M.) 199–213 (1972).
14. Dini, J. *Electrodeposition* (Noyes, New Jersey, 1992).

15. Ramasany, P. *Handbook of Crystal Growth* 1a (ed. Hurler, C. J.) (North Holland, Amsterdam, 1993).
16. Bergstrasser, T. R. & Merchant, H. D. in *Defect Structure, Morphology and Properties of Deposits* (ed. Merchant, H. D.) Proceedings of the Materials Week, Rosemont, 115–168 (TMS publications, Warrendale, PA, USA).
17. Vetter, K. J. *Electrochemical Kinetics—Theoretical Aspects* Ch. 2, section D, 282–327 (Academic, New York, 1967).
18. Budevski, E., Staikov, G. & Lorenz, W. J. *Electrochemical Phase Formation and Growth, an Introduction to the Initial Stages of Metal Deposition* Chs 4, 5 (VCH, New York, 1996).

Acknowledgements. I am grateful for discussions with D. Barkey; the cell design was in part elaborated in collaboration with him. One series of experiments was prepared with N. Lidgi. I thank A. Chauvineau for lending the AFM. I also shared many stimulating discussions with J.-F. Gouyet, M. Plapp, B. Sapoval, J.-N. Chazalviel, M. Rosso and C. Brissot of the Laboratoire PMC. I acknowledge the help of L.-A. Courrière with image processing.

Correspondence should be addressed to the author (e-mail: vincent.fleury@polytechnique.fr).

Discovery of a second family of bismuth-oxide-based superconductors

S. M. Kazakov[†], C. Chaillout^{*}, P. Bordet^{*}, J. J. Capponi^{*}, M. Nunez-Regueiro[‡], A. Rysak[§], J. L. Tholence[§], P. G. Radaelli^{||}, S. N. Putilin[†] & E. V. Antipov[†]

^{*} Laboratoire de Cristallographie, CNRS, BP 166, 38042 Grenoble cedex 9, France

[†] Chemistry Department, Moscow State University, Moscow 119899, Russia

[‡] EPM-Matformag, CNRS, BP 166, 38042 Grenoble cedex 9, France

[§] L.E.P.E.S., CNRS, BP 166, 38042 Grenoble cedex 9, France

^{||} Institut Laue–Langevin, BP 156, 38042 Grenoble cedex 9, France

The superconducting oxide $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_3$, discovered in 1975 (ref. 1), is an exotic system having an unusually high transition temperature (T_c) of ~ 12 K, despite a relatively low density of states at the Fermi level. The subsequent prediction² that doping the electronically inactive barium donor sites, instead of the bismuth sites, might induce superconductivity with a higher T_c led to the discovery^{3,4} in 1988 of superconductivity in the $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ system ($T_c \sim 30$ K for $x = 0.4$). But it remains an open question why many of the superconducting properties of these materials are similar to those of the well-known copper oxide superconductors⁵, despite their pronounced structural differences: the former have a three-dimensional bismuth–oxygen framework, whereas the structures of the latter are predominantly two-dimensional, consisting of copper–oxygen planes. Understanding of the copper oxide superconductors has gained immensely from the study of many different superconducting systems, and so it might be expected that the identification of bismuth oxide superconductors beyond the substituted BaBiO_3 compounds will prove to be similarly fruitful. Here we report the synthesis of a second family of superconducting bismuth oxides, based on SrBiO_3 . We show that partial substitution of potassium or rubidium for strontium induces superconductivity with T_c values of ~ 12 K for $\text{Sr}_{1-x}\text{K}_x\text{BiO}_3$ ($x = 0.45–0.6$) and ~ 13 K for $\text{Sr}_{1-x}\text{Rb}_x\text{BiO}_3$ ($x = 0.5$).

The first step of the synthesis of samples consisted of the preparation of $\text{Sr}_2\text{Bi}_2\text{O}_5$ by annealing the appropriate mixture of SrCO_3 and Bi_2O_3 in air at 850°C for 50 h, with two intermediate regrindings. SrBiO_3 was then obtained by heat-treating $\text{Sr}_2\text{Bi}_2\text{O}_5$ at 800°C for 35 h under high oxygen pressure (100 bar), which was necessary to stabilize this compound. In contrast, the K-doped samples, with a nominal $\text{K}/(\text{K} + \text{Sr})$ ratio between 0.2 and 0.8, were prepared by a high-pressure–high-temperature technique in a belt-type apparatus (2 GPa, 700°C , 1 h, Pt capsules) with stoichiometric mixtures of $\text{Sr}_2\text{Bi}_2\text{O}_5$, Bi_2O_3 and KO_2 .

The samples were first characterized by X-ray diffraction. Those with a K content in the range $x = 0.45–0.6$ appear as a single perovskite-like phase, whereas the samples with a lower K content contained traces of bismuth-based oxides as impurities. The cationic composition of the different samples was also checked by energy-

dispersive spectroscopy (EDS) in a transmission electron microscope. We found that the solubility limit of K is $\sim 60\%$ in our synthesis conditions. However, it might depend on the oxygen stoichiometry of the initial mixture.

From a.c. susceptibility and resistivity measurements, superconductivity occurs for samples with $0.45 \leq x \leq 0.60$ with a T_c of the order of 12 K. Figure 1 gives an example of the temperature dependence of the resistance and the a.c. susceptibility for a sample with $x = 0.6$. The superconducting volume fraction is determined from a.c. susceptibility measurements (at 0.1 Oe and 119 Hz) on finely ground powders (instead of ceramics, to avoid overestimating the superconducting volume), by comparison with $-3/8\pi$, the susceptibility of spherical grains. The superconducting volume at 1.5 K is $\sim 40\%$ for $x \approx 0.6$, confirming the bulk nature of superconductivity. The superconducting volume and temperature do not change for the samples with a K content between 0.55 and 0.6, whereas materials with lower K contents (0.45–0.55) exhibit a smaller superconducting fraction ($\leq 10\%$) with $T_c = 10$ –11 K.

All X-ray diffractograms, as well as electron diffraction patterns, can be indexed on the basis of a distorted perovskite cell with lattice parameters $a \approx b \approx \sqrt{2}a_p$, $c \approx 2a_p$, where a_p refers to the ideal cubic perovskite cell ($a_p \approx 4$ Å). For SrBiO_3 , the lattice is metrically very close to orthorhombic, with possibly a small monoclinic distortion. The electron diffraction pattern (Fig. 2a) shows that the crystallites

are twinned as expected from the close values of a , b and $c/\sqrt{2}$. As a result of K doping the lattice parameters decrease, as does the distortion from cubic symmetry. In the superconducting region the symmetry appears tetragonal, as indicated by X-ray powder diffraction. Furthermore, from the electron diffraction pattern, the lattice is body-centred (Fig. 2b). For $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_3$ and $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ the symmetry also changes as a function of doping and superconductivity occurs close to a phase transition leading to either tetragonal⁶ or cubic⁷ symmetry, respectively.

To precisely determine the symmetry and the structure of $\text{Sr}_{1-x}\text{K}_x\text{BiO}_3$, we collected neutron diffraction data on the D2B diffractometer at the Institut Laue–Langevin ($\lambda = 1.59432$ Å) on two powdered samples with $x = 0$ and $x = 0.6$ (from EDS analysis). The structural refinements were made with the GSAS program⁸.

For SrBiO_3 , a certain ambiguity remains between the orthorhombic $Pnma$ and the monoclinic $P2_1/n$ space groups, as satisfactory refinements can be obtained in both cases, and the refined monoclinic angle ($90.063(6)^\circ$) is rather small (the metric distortion is about half that of BaBiO_3 (ref. 9)). However, the monoclinic $P2_1/n$ space group yields a significantly lower χ^2 (1.60 compared with 1.85, with only four additional structural parameters) and U , V , W peak profile parameters very close to the D2B instrument's resolution, whereas a significant broadening needed to be introduced for the $Pnma$ model. In the $P2_1/n$ space group, the Bi cations occupy two

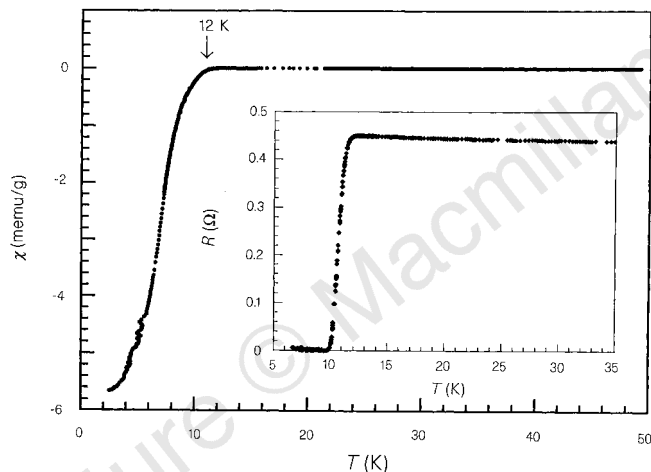


Figure 1 Temperature dependence of the a.c. magnetic susceptibility, χ , and the resistance (inset) of a $\text{Sr}_{0.4}\text{K}_{0.6}\text{BiO}_3$ sample.

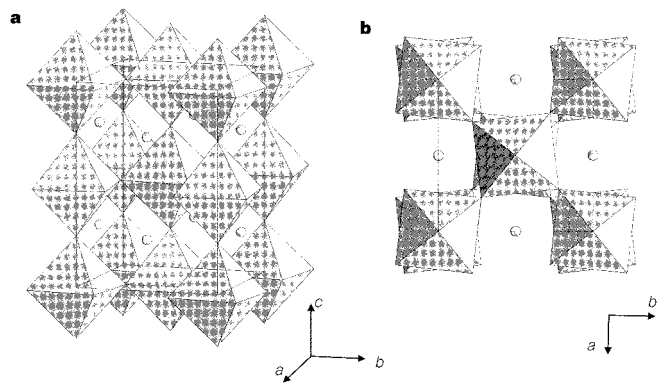


Figure 3 Structural model of the superconducting compound $\text{Sr}_{0.4}\text{K}_{0.6}\text{BiO}_3$. **a**, Structure; the Sr(K) cations are represented as spheres. **b**, Projection of the structure along the c -axis, showing the tilts of the BiO_6 octahedra.

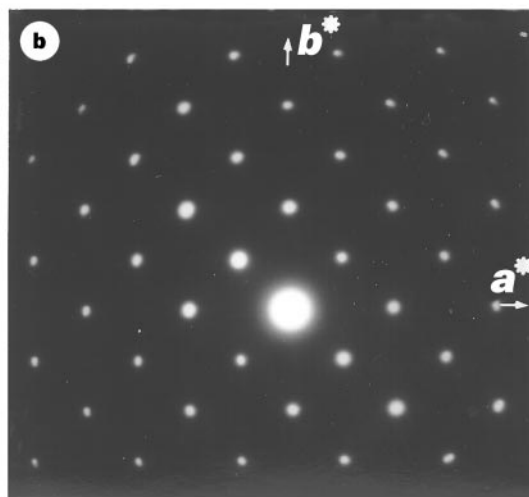
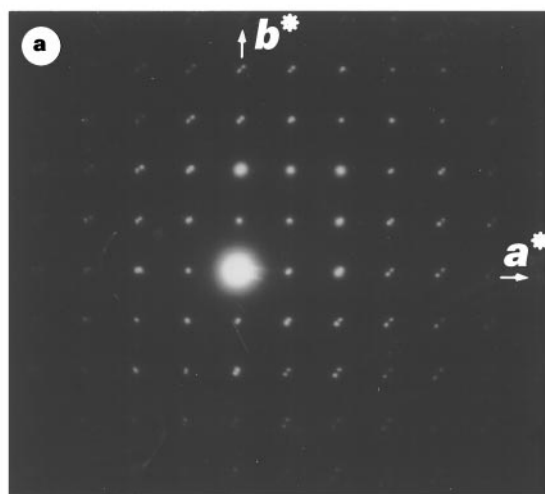


Figure 2 Electron diffraction patterns taken along the $[001]$ zone axis for crystallites of SrBiO_3 (**a**; the splitting of the reflections indicates that the crystallite is twinned) and $\text{Sr}_{0.4}\text{K}_{0.6}\text{BiO}_3$ (**b**).

Table 1 Positional and thermal parameters, selected interatomic distances and average Bi valence for SrBiO₃ and Sr_{0.4}K_{0.6}BiO₃

	SrBiO ₃	Sr _{0.4} K _{0.6} BiO ₃
Space group	<i>P</i> 2 ₁ / <i>n</i>	<i>I</i> 4/ <i>mcm</i>
Lattice parameters	<i>a</i> = 5.9480(2), <i>b</i> = 6.0951(2), <i>c</i> = 8.4854(3) Å, β = 90.063(6)°	<i>a</i> = 5.9416(2) Å, <i>c</i> = 8.4394(4) Å
Sr(K) <i>x</i> , <i>y</i> , <i>z</i>	−0.0152(7), 0.5455(4), 0.251(1)	0, 0.5, 0.75
<i>U</i> (Å ²)	0.0078(7)	0.016(1)
Bi1 <i>x</i> , <i>y</i> , <i>z</i>	0, 0, 0	0, 0, 0
<i>U</i> (Å ²)	0.0045(6)	0.0017(5)
Bi2 <i>x</i> , <i>y</i> , <i>z</i>	0, 0, 0.5	—
<i>U</i> (Å ²)	0.0045(6)	—
O1 <i>x</i> , <i>y</i> , <i>z</i>	0.4056(7), 0.4617(7), 0.241(1)	0, 0, 0.25
<i>U</i> (Å ²)	0.019(1)	<i>U</i> 11, <i>U</i> 33: 0.033(3), 0.009(2)
O2 <i>x</i> , <i>y</i> , <i>z</i>	0.284(1), 0.190(1), 0.546(1)	0.2219(5), 0.7219(5), 0
<i>U</i> (Å ²)	0.019(2)	<i>U</i> 11, <i>U</i> 33: 0.011(1), 0.060(3)
O3 <i>y</i> , <i>y</i> , <i>z</i>	0.187(1), 0.714(1), 0.556(1)	—
<i>U</i> (Å ²)	0.012(2)	—
<i>R</i> _w , <i>χ</i> ²	7.11%, 1.60	8.19%, 1.88
Sr(K)–O (Å)	3.484(6), 2.556(6), 3.618(5), 2.620(5), 3.76(1), 3.01(1), 2.85(1), 2.55(1), 3.03(1), 3.83(1), 2.50(1), 2.83(1)	2.9708(2) × 4, 2.814(2) × 4, 3.150(2) × 4
Bi1–O (Å)	2.28(1) × 2, 2.317(7) × 2, 2.321(6) × 2	2.1098(4) × 2, 2.114(3) × 4
Bi2–O (Å)	2.14(1) × 2, 2.088(7) × 2, 2.123(7) × 2	—
<i>v</i> (Bi1)	3.21	4.62
<i>v</i> (Bi2)	4.57	—

U, mean square temperature displacement; *R*_w, weighted profile agreement factor; *v*, valence.

distinct distorted octahedral sites. The positional and thermal parameters after the last cycle of refinement, as well as the Bi–O and Sr–O interatomic distances, are given in Table 1. Attempts to refine the cationic or anionic site occupancies did not lead to any deviation from full occupancies. The average Bi–O distances are 2.30(1) and 2.12(1) Å for the two sites. The average valences of Bi1 and Bi2 calculated with the Zachariasen formula¹⁰ and the constants reported previously¹¹ are +3.21 and +4.57, respectively. As with BaBiO₃^{9,11}, this indicates a partial disproportionation of the Bi cations into Bi³⁺ and Bi⁵⁺, which occupy the two distinct crystallographic Bi1 and Bi2 sites. This charge localization is in agreement with the insulating properties of this compound.

The structure of Sr_{0.4}K_{0.6}BiO₃ has been refined in the *I*4/*mcm* space group (as with BaPb_{1–x}Bi_xO₃ in the superconducting region). The main results are reported in Table 1 and a structural model is shown in Fig. 3. The refined K content is 0.56(2), in good agreement with the EDS data. The temperature factors of the oxygen atoms are highly anisotropic with anomalously large components. When refined, the occupancy of the oxygen sites was close to unity. The anomalous values of the temperature factors might reflect a positional disorder of the oxygen atoms depending on whether the neighbouring cations are K or Sr. The Bi cations occupy only one octahedral site with Bi–O distances 2.1098(4) × 2 and 2.114(3) × 4 Å. This yields an average Bi valence of +4.62, which is in good agreement with the value expected from the refined composition (+4.56).

Our preliminary study of the Rb-substituted samples, Sr_{1–x}Rb_xBiO₃, shows that these materials, prepared in the same way as the K-substituted ones but using Rb₂O instead of KO₂, exhibit superconductivity with *T*_c = 13 K at *x* ≈ 0.5. Owing to the small superconducting volume fraction (~6%) and poor phase purity for Rb-substituted samples, further investigations are necessary to clarify the origin of the superconductivity in this system.

A comparison of the systems Sr_{1–x}K_xBiO₃, BaPb_{1–x}Bi_xO₃, Ba_{1–x}K_xBiO₃ and BaPb_{1–x}Sb_xO₃ (ref. 12), with maximum *T*_c values of ~12 K, 12 K, 30 K and 3.5 K respectively, has shown that keeping the perovskite B site intact is not the only parameter that provides a high *T*_c. The octahedral tilt scheme influenced by the size of the A cation is also an important aspect. This new system indicates that, in contrast with the Cu-based superconductors where the two-dimensional character seems necessary for a high *T*_c, the highest three-dimensional symmetry seems to be the clue for a high *T*_c in bismuth oxides. □

Received 11 April; accepted 3 September 1997.

1. Sleight, A. W., Gillson, J. L. & Bierstedt, P. E. High-temperature superconductivity in the BaPb_{1–x}Bi_xO₃ system. *Solid State Commun.* **17**, 27–28 (1975).

- Mattheiss, L. R., Gyorgy, E. M. & Johnson, D. W. Superconductivity above 20 K in the Ba–K–Bi–O system. *Phys. Rev. B* **37**, 3745–3746 (1988).
- Cava, R. J. *et al.* Superconductivity near 30 K without copper: the Ba_{0.4}K_{0.6}BiO₃ perovskite. *Nature* **332**, 814–816 (1988).
- Hinks, D. G. *et al.* Synthesis, structure and superconductivity in the Ba_{1–x}K_xBiO_{3–y} system. *Nature* **333**, 836–838 (1988).
- Klein, T., Harneit, W., Baril, L. & Escribe-Filippini, C. Pinning and vortex dynamics in superconducting (K,Ba)BiO₃ single crystals. *J. Low Temp. Phys.* **105**, 1067–1073 (1996).
- Marx, D. T. *et al.* Metastable behavior of the superconducting phase in the BaPb_{1–x}Bi_xO₃ system. *Phys. Rev. B* **46**, 1144–1156 (1992).
- Pei, S. *et al.* Structural phase diagram of the Ba_{1–x}K_xBiO₃ system. *Phys. Rev. B* **41**, 4126–4141 (1990).
- Larson, A. C. & von Dreele, R. B. Report LA-UR-86-748 (Los Alamos Laboratory, Univ. California, 1985).
- Cox, D. E. & Sleight, A. W. Crystal structure of Ba₂Bi³⁺Bi⁵⁺O₆. *Solid State Commun.* **19**, 969–973 (1976).
- Zachariasen, W. H. & Penneman, R. A. Application of bond length–strength analysis to 5f element fluorides. *J. Less Common Metals* **69**, 369–377 (1980).
- Chaillout, C. *et al.* Bismuth valence order–disorder study in BaBiO₃ by powder neutron diffraction. *Solid State Commun.* **65**, 1363–1369 (1988).
- Cava, R. J. *et al.* Superconductivity at 3.5 K in BaPb_{0.75}Sb_{0.25}O₃: why *T*_c so low? *Nature* **339**, 291–293 (1989).

Acknowledgements. We thank M. Perroux for his help in the high-pressure experiments. This work was partly supported by the Nato Linkage, a French–Russian network of the French MENSUR and the Russian project 'POISK'.

Correspondence and requests for materials should be addressed to C.C. (e-mail: chaillout@polycnrs-gre.fr).

Chemical composition of dissolved organic nitrogen in the ocean

Matthew McCarthy*, Tom Pratum†, John Hedges* & Ronald Benner‡

* School of Oceanography, Box 357940, and † Department of Chemistry, Box 351700, University of Washington, Seattle, Washington 98195-7940, USA
‡ University of Texas, Marine Science Institute, Port Aransas, Texas 78373, USA

Fixed nitrogen is one of the main limiting nutrients for primary production in the ocean^{1–3}, where it is biologically available in the form of dissolved inorganic and organic matter. Inorganic nitrogen concentrations are consequently very low in surface waters of temperate ocean gyres, yet fixed nitrogen persists in the form of dissolved organic matter. The small, rapidly cycling organic compounds fundamental to microbial and planktonic growth (such as free amino acids, amines and urea^{4,5}) account for only a minor fraction of total dissolved organic nitrogen (DON). In contrast, the vast majority of DON, especially in the deep ocean, resides in the form of nitrogenous substances that are resistant to biological degradation. These substances, which represent an

enormous reservoir of fixed nitrogen, are not readily identified by conventional biochemical techniques, but have been assumed to consist largely of structurally complex macromolecules resulting from the degradation and spontaneous abiotic condensation of biochemical precursors⁶. Here we present ¹⁵N NMR measurements that contradict this view. Our results show that most higher-molecular-weight DON in the ocean exists in amide form, rather than as a collection of nitrogen heterocycles that might be indicative of spontaneous condensation products. Because these amides are unlikely to form abiotically, the bulk of the ocean's DON reservoir appears to derive directly from degradation-resistant biomolecules.

Ultrafiltered dissolved organic matter (UDOM) was isolated from four depths (surface to 4,000 m depth) in the central Equatorial Pacific Ocean, and from surface waters along a transect in the oligotrophic central North Pacific gyre. These samples include both biologically active surface waters and less dynamic abyssal waters from ocean regions free of direct terrestrial influence. Samples were isolated using tangential-flow ultrafiltration, a size-based technique that reproducibly isolates 20–30% of total DOM from sea water^{7,8}. Retention of total DOM ranged from 19% in the deep Pacific to 38% in surface waters, with atomic C/N ratios of the isolates ranging from 15.6 to 18.4 (Table 1).

UDOM represents the larger size fraction of the operationally defined (<0.1 µm) DOM pool, and the degree to which it is compositionally representative of lower-molecular-weight oceanic DOM remains to be fully determined. Continuous enzymatic degradation of large to smaller molecules takes place in the water column, and all dissolved material surviving over longer timescales is likely to be affected by this process. Bulk properties of UDOM, including C/N ratios, stable carbon isotopic ratios, radiocarbon 'ages' and overall amino-acid content, are similar to the same properties for total DOM, and are clearly distinct from those of marine biomass or detrital particles^{8–10}. Although it is likely that some compositional differences within molecular weight ranges do exist, the comparisons of C/N ratio and overall amino-acid content suggest that UDOM isolation does not greatly fractionate DOM in terms of major nitrogenous components.

Cross-polarization magic-angle-spinning (CPMAS) techniques in nuclear magnetic resonance spectroscopy of solid samples provide a powerful tool for elucidating structural characteristics of complex natural organic mixtures. However, because the low natural abundance and small magnetogyric ratio of the ¹⁵N nucleus results in low sensitivity, ¹⁵N CPMAS NMR has usually been used only on isotopically enriched materials¹¹. Recently, with the use of higher magnetic fields and long signal averaging times, ¹⁵N CPMAS NMR has been successfully applied in studies of natural organic matter in soils¹², coal¹³ and marine sedimentary deposits¹⁴. Here we report application of ¹⁵N CPMAS NMR with large DOM samples available using ultrafiltration, providing the first comprehensive view of natural nitrogen forms in this major ocean reservoir.

The ¹⁵N CPMAS spectra of UDOM samples isolated from the central Pacific Ocean are characterized by a single well-defined peak centred near 260 p.p.m. (Fig. 1a). Although the peak width at half-maximum of the 260 p.p.m. band is similar in surface samples, the width of the same resonance increases in deep waters. This resonance corresponds mainly to the chemical shift range for amide nitrogen¹⁵. Comparison with similarly obtained ¹⁵N CPMAS spectra of several reference biochemicals (Fig. 1b) confirms the assignment: two of the most common amide-containing biopolymers, protein and chitin, have ¹⁵N resonances at 259–261 p.p.m. (Fig. 1b) and cannot be distinguished at the resolution of our solid-state NMR experiment. As well as verifying the chemical shift assignments, the reference spectra demonstrate our ability to observe diverse nitrogen species (amides, amides and heterocyclic nitrogen). Apart from the main peak, no other signals can be unambiguously identified in any sample, indicating that only negligible amounts of more-

shielded nitrogen (for example, aliphatic amines and non-acylated amino sugars) or less-shielded nitrogen (for example, heterocyclic or imine) are present.

The chlorophyll maximum sample from 100 m is characterized by the higher overall DOC, the lowest C/N and shifts in ¹⁵N stable isotope values indicative of active production⁸. The ¹⁵N NMR spectrum of this sample also displays the narrowest and most symmetrical resonance, with a chemical shift (260 p.p.m.), width

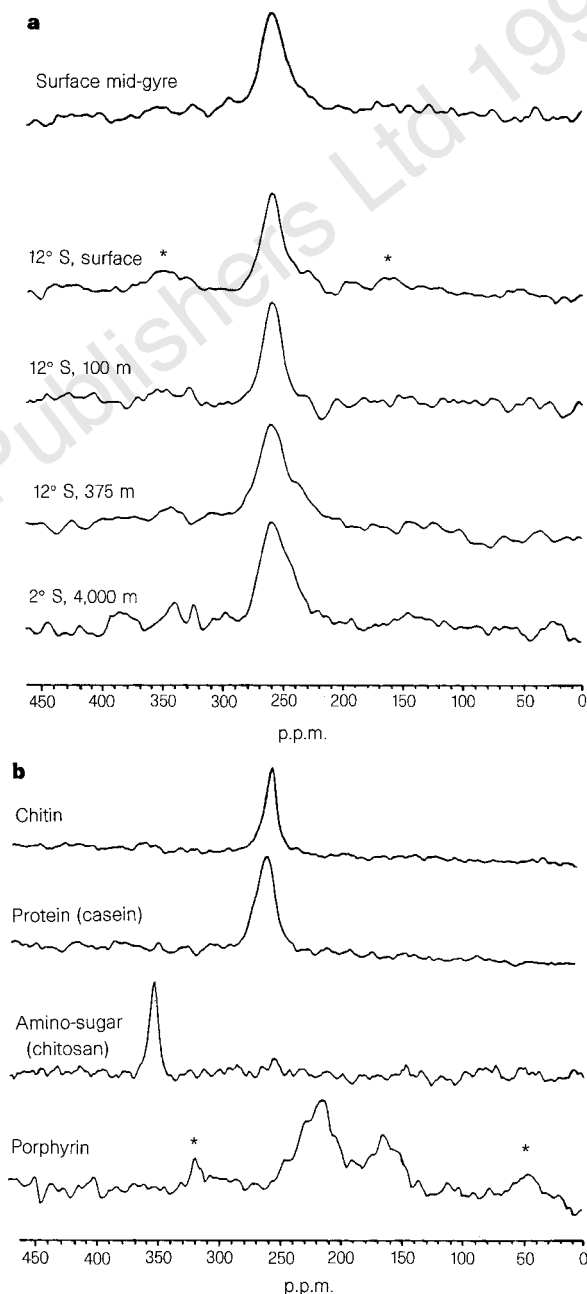


Figure 1 Natural abundance ¹⁵N CPMAS spectra for Pacific UDOM samples and selected natural materials. Spinning sidebands (where clearly observable) are indicated by asterisks. **a**, Spectra for Pacific Ocean UDOM samples. The 12° S surface sample was acquired at a slower spin rate of 3.0 kHz, resulting in spinning sidebands for the amide peak. A shortened pulse delay of 500 ms was used to acquire roughly 500,000 transients for the 12° S 375 m and surface samples after replicate analyses of other samples at 500 ms delay resulted only in improved signal-to-noise ratios. **b**, Natural abundance spectra for selected nitrogen-containing biochemicals. Amides: chitin (*N*-acetylglucosamine polymer) and casein (a protein). Amino sugar: chitosan (deacetylated chitin). Heterocyclic nitrogen: protoporphyrin IV.

at half-maximum (20 p.p.m.), and range (237–277 p.p.m.) similar to those of the pure amide biopolymers (Fig. 1b). This observation suggests that nearly all UDOM nitrogen at this depth is present as biosynthesized amides.

Because 60–80% of organic nitrogen in biomass or marine particles is recoverable as hydrolysable amino acids¹⁶, peptides might be expected to predominate in UDOM from upper waters. However, only a minor percentage (<10%) of total nitrogen could be recovered from any UDOM isolate as hydrolysable amino acids (Table 1). Furthermore, this fraction of UDOM nitrogen showed little variation with depth, being 8–10% in all samples. The absence of elevated values in surface samples is surprising, indicating that freshly produced UDOM is not greatly enriched in hydrolysable peptide. Similarly low yields have been observed in depth profiles from other ocean basins⁹. The persistence of hydrolysable peptide in the deep ocean suggests that a small fraction of biomass protein—the likely source of this material—is preserved over long timescales. In fact, specific membrane proteins have been found to persist intact in both the surface and deep ocean¹⁷, suggesting that some proteinaceous materials are either difficult to degrade or physically protected from enzymatic hydrolysis.

The unknown amide fraction makes up the majority of total nitrogen in all UDOM samples. This amide could be proteinaceous material which has been rendered largely resistant to conventional hydrolysis methods. However, although there is evidence that acid hydrolysis does not fully account for hydrolysable amino acids in oceanic DOM¹⁸, even a several-fold increase in yield would leave the majority of amide unaccounted for. In addition, the previous⁷ noted lack of even a small elevation in amino-acid yield in upper water UDOM does not support the explanation of a proteinaceous material that has become physically protected over time. Curie-point pyrolysis of previous Pacific UDOM samples¹⁹ has also yielded low abundances of amino-acid-derived products. Thus, although some hydrolysis-resistant peptide is likely to be present, other amide forms may also be important components of UDOM nitrogen.

If a substantial portion of the unknown amide is not proteinaceous, the high concentrations and fluxes imply another common and abundant biochemical source. Unusual hydrolysis-resistant amide structures have recently been indicated in some algae²⁰, but have rarely been reported and are likely to be present only in particulate form. Acetylated amino sugars are the second major

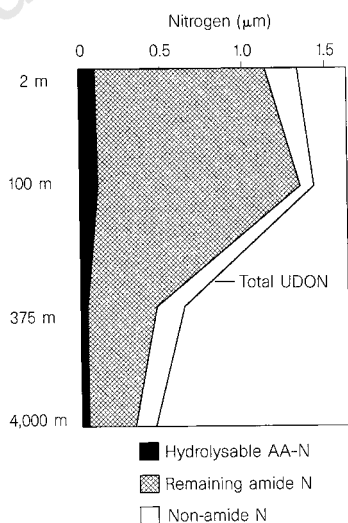


Figure 2 Organic UDOM nitrogen forms in Pacific Ocean. Distribution of the three broad categories of ultrafiltered DON from the depth profile at 12°S, 140°W, and deep-water sample (4,000 m) at 2°S, 140°W in the Pacific Ocean. As described in the text, the differences between amide N and total UDOM concentrations are probably made up by stable heterocyclic nitrogen forms (pyrrole/indole-like nitrogen). Note that the depths on the y-axis are not to scale.

Table 1 Pacific UDOM recovery and bulk ON composition

	2 m (mid-gyre)	2 m (surface)	100 m (chl. max.)	375 m (O ₂ min.)	4,000 m (deep)
DOC (μM)	78	82	85	53	45
% UDOM	30	38	27	19	22
(C/N)	16.4	16.8	15.6	16.9	18.4
UDON (μM)	1.12	1.33	1.42	0.63	0.45
%AA-N	n.d.	8.2	8.2	8.5	9.4
% other amide	75	77	86	66	65
% non-amide	17	15	6	25	26
Centre (p.p.m.)	260	259	261	261	261
Width (p.p.m.)	23	21	20	27	32
Total range (p.p.m.)	217–287	225–285	237–277	206–283	222–280

Total dissolved organic carbon (DOC) values, as well as of carbon-based ultrafiltered dissolved organic matter recoveries (%UDOM), for the central Pacific water samples (12°S, 140°W), including depths corresponding to the oxygen minimum (O₂ min) and chlorophyll maximum (chl. max.). Total UDOM concentrations are derived from CHN data, the dry weight recovered and the volume filtered. For UDOM isolates, we give atomic carbon/nitrogen ratios (C/N) and estimates of the contribution made by the three nitrogen classes discussed to total UDOM at each depth. Percentage amino-acid nitrogen (%AA-N) is calculated directly from hydrolysable amino-acid data, while the percentage non-amino-acid amide (% other-amide) and the percentage non-amide are derived from integration of the ¹⁵N NMR spectra, based on resonance ranges of authentic amide biopolymers. For ¹⁵N CPMAS NMR spectra the centre resonance, full width at half maximum and total range at baseline are also given.

amide-containing constituent of biomacromolecules, occurring widely in structural polymers of plants, animals and bacteria²¹. Common marine sources include chitin, a structural component in many eukaryotic plankton species, and bacterial cell wall components, such as mureins and lipopolysaccharides. The potential for long-term preservation of such structural polymers is generally thought to be greater than that for proteins or storage carbohydrates²¹, and peptidoglycans in particular contain many uncommon constituents that probably act to deter common hydrolytic enzymes²². UDOM from both surface and deep waters is enriched in carbohydrate characterized by low neutral sugar content and constant molecular-level composition⁹, suggesting that a similar range of structural polysaccharides may be ubiquitous throughout the oceanic water column^{9,23}. Although hydrolysis and quantification of charged sugar residues is difficult²⁴, pyrolysis of UDOM samples has also indicated *N*-acetyl amino sugar precursors suggestive of structural polymers¹⁹.

Although amide nitrogen clearly dominates all samples, there are also indications of smaller contributions by other nitrogen forms, especially in samples from deep waters with low overall UDOM concentration. Widths of the 260 p.p.m. peak at half maximum, a measure relatively unaffected by baseline noise, clearly increases with water depth (Table 1). The increase is not symmetrical, but instead falls mostly in a small downfield shoulder between 220 and 240 p.p.m. This spectral region is outside the range of authentic amide biopolymers, and corresponds mainly to indole- and pyrrole-like nitrogen¹⁵. Such heterocyclic structures are among the most stable organic nitrogen forms in the geosphere, accounting for much of the organic nitrogen observed in petroleum²⁵, coals¹³ and some marine sediments^{14,25}. Natural marine sources of 'pyrrole-like' nitrogen resonating in the same region include the bases of DNA and RNA (purines and pyrimidines) and the porphyrin backbone of photosynthetic pigments. Although these compounds are present in the ocean at low concentrations, their bases may be relatively stable. In addition to biological sources, it is possible that compounds formed abiotically also contribute. Melanoidins, the complex end-products of sugar and amino-acid condensation, also resonate in this region²⁶, and their formation has been long hypothesized as an important abiotic pathway for organic matter preservation²⁷. Although it is not possible to distinguish specific sources for this signal, the 220–240 p.p.m. shoulder does appear as a consistent feature of all spectra.

Taken together, the ^{15}N NMR, amino acid, and total UDOM nitrogen concentrations in the central Pacific indicate oceanographically consistent depth distributions of three principal nitrogen types throughout the oligotrophic ocean (Table 1, Fig. 2). Although the reactivity of these three classes through the water column seems to be generally similar, the data do suggest some distinct cycling. The downfield shoulder suggests that a small portion of total UDOM nitrogen at all depths is non-amide. This fraction has its highest relative abundance (26%) in the deep ocean, consistent with NMR evidence that it is made up of stable—and thus slowly cycling—heterocyclic nitrogen forms (pyrrole/indole-like nitrogen). Hydrolysable peptide makes up a surprisingly small fraction, accounting for only about 10% of total UDOM nitrogen at all depths. It is the remaining amide component, mainly as a result of its high abundance, that drives almost all of the water-column gradient in UDOM nitrogen. This observation suggests that hydrolysis-resistant peptide or other amide-containing biomolecules may fuel most DON remineralization below the euphotic zone, a process which (coupled with advection) acts to close overall nitrogen budgets in the upper ocean².

Despite these gradients, surface-produced amide also constitutes the main component of UDOM nitrogen in the deep Pacific, suggesting that traditional paradigms for DOM cycling in the ocean may need to be re-evaluated. Over 80% of the organic nitrogen in the ocean resides at low concentration in the enormous subsurface reservoir, an environment in which few of the factors commonly understood to control organic matter preservation, such as anoxia, burial or protection by mineral surfaces, can be invoked. This deep-ocean material has overall radiocarbon 'ages' of about 4,000–6,000 years (ref. 28), an observation which has reinforced long-standing hypotheses that condensation reactions produce chemically altered geopolymers which, by virtue of unusual chemical structure, persist over many ocean mixing cycles⁶. Recent work has indicated that specific polysaccharides^{9,23} as well as proteins¹⁷ may survive in the ocean over long timescales. However, the wholesale persistence of amide functional groups within the larger-molecular-weight fraction of abyssal DOM indicates that traditional geopolymerization is not a dominant preservation mechanism for this material. In addition, the overall similarities of ^{14}C 'ages', C/N ratios and amino-acid composition between UDOM and total DOM suggest that most organic nitrogen in the ocean retains the principal structural features of the parent biomolecules. Identifying the processes that render common biochemical structures resistant to enzymatic degradation will be central to understanding controls for storage of reduced carbon in the deep sea. □

Methods

Nitrogen-15 CPMAS NMR spectra were obtained on a Bruker AF-300 spectrometer at a frequency of 30.4 MHz and a spin rate of 3.5 kHz. Chemical shifts are referenced to external NO_3^- . Spectral parameters were chosen based on those recently determined as optimum for quantitative ^{15}N solid-state spectra in a variety of natural materials^{6,8}. For most spectra, a contact time of 1 ms, an acquisition time of 20 ms and a pulse delay of 1.0 s were used to acquire roughly 250,000 transients. Before Fourier transformation, 125 Hz of line broadening was applied. Neutron activation analysis of a separate Pacific profile (0–4,000 m) indicates that total concentrations of paramagnetic metals in central Pacific UDOM are low (<0.01%) and constant with depth. Amide/non-amide regions were estimates based on resonance ranges of authentic amide biopolymers, and were integrated using Lab-Calcul NMR software. The variability of peak integration due to noise is less than 10%, based on integration of equivalent noise regions over the spectral range considered.

Previous work has indicated that optimum acquisition parameters for natural organic mixtures are broad and similar^{12,13,29}. Several factors, however, may reduce signal intensities. Although the contact time used might be too short for full cross-polarization of non-protonated N, increasing the contact time could result in attenuation from rotating frame relaxation effects. Quaternary carbon is readily observable in ^{13}C CPMAS UDOM spectra using

the same cross-polarization conditions^{7,11}, and previous studies have indicated that these contact times are adequate for observation of unsaturated heterocyclic nitrogen^{13,30}. Nitrogen in asymmetric electronic environments, such as unsaturated or aromatic nitrogen, can also have intensity diverted into spinning sidebands as a result of anisotropic chemical shift effects. The porphyrin in Fig. 1b shows that the unsaturated heterocyclic nitrogens, although observable, are somewhat attenuated by sidebands. Thus, the detection limit for unsaturated heterocyclic nitrogen is likely to be higher than for other nitrogen forms.

Isolation of UDOM using tangential flow ultrafiltration, measurement of total dissolved organic carbon and determination of UDOM recoveries were carried out as described previously^{8,9,31}. Total carbon and nitrogen contents of dried samples were measured after vapour-phase acidification with a precision of approximately $\pm 2\%$ using a Carlo Erba 1108 CHN analyser³². Hydrolysable amino acids were determined using charge-matched recovery standards after 6 M HCl hydrolysis¹⁶, with individual amino acids quantified as pentafluoropropyl isopropyl esters by gas chromatography with flame ionization detection. Analytical precision for this procedure is $\pm 10\%$ or less for total amino acid yields. The percentage of amino-acid nitrogen (%AA-N) was calculated from the sum of amino-acid nitrogen and the total elemental nitrogen content¹⁶.

Received 7 October 1996; accepted 5 September 1997.

1. Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* (ed. Hill, N. H.) 26–76 (Wiley Interscience, New York, 1963).
2. Smith, S. V., Kimmmer, W. J. & Walsh, T. W. Ventral flux and biogeochemical turnover regulate nutrient limitation of net organic production in the North Pacific Gyre. *Limnol. Oceanogr.* **31**, 161–167 (1986).
3. Jackson, G. A. & Williams, P. M. Importance of dissolved organic nitrogen and phosphorus to biological nutrient cycling. *Deep Sea Res.* **32**, 223–235 (1985).
4. Bronk, D. A. & Glibert, P. M. Nitrogen uptake, dissolved organic nitrogen release, and new production. *Science* **265**, 1843–1852 (1994).
5. Keil, R. G. & Kirchman, D. L. Contribution of dissolved free amino acids and ammonium to the nitrogen requirements of heterotrophic bacterioplankton. *Mar. Ecol. Prog. Ser.* **73**, 1–10 (1991).
6. Harvey, G. R., Boran, D. A., Chesal, L. A. & Tokar, J. M. The structure of marine fulvic and humic acids. *Mar. Chem.* **12**, 119–132 (1983).
7. Benner, R., Pakulski, J. D., McCarthy, M., Hedges, J. I. & Hatcher, P. G. Bulk chemical characteristics of dissolved organic matter in the ocean. *Science* **255**, 1561–1564 (1992).
8. Benner, R., Biddanda, B., Black, B. & McCarthy, M. Abundance, size distribution, and stable carbon and nitrogen isotope compositions of marine organic matter isolated by tangential-flow ultrafiltration. *Mar. Chem.* **57**, 243–263 (1997).
9. McCarthy, M. D., Hedges, J. I. & Benner, R. Major biochemical composition of dissolved high-molecular weight organic matter in seawater. *Mar. Chem.* **55**, 281–297 (1996).
10. McCarthy, M. D., Hedges, J. I. & Benner, R. The chemical composition of dissolved organic matter in seawater. *Chem. Geol.* **107**, 503–507 (1993).
11. Preston, C. M. Applications of NMR to soil organic matter analysis: history and prospects. *Soil Sci.* **161**, 144–166 (1996).
12. Knicker, H., Frund, R. & Ludermann, H. D. The chemical nature of nitrogen in native soil organic matter. *Naturwissenschaften* **80**, 219–221 (1993).
13. Knicker, H., Hatcher, P. G. & Scaroni, A. W. Solid-State ^{15}N NMR spectroscopy of coal. *Energy Fuels* **9**, 999–1102 (1995).
14. Knicker, H., Hatcher, P. G. & Scaroni, A. W. ^{13}C and ^{15}N -NMR spectroscopic investigation on the formation of fossil algal residues. *Org. Geochem.* **67**, 661–669 (1996).
15. Levey, G. C. & Lichter, R. L. *Nitrogen-15 NMR Spectroscopy* (Wiley, New York, 1979).
16. Cowie, G. L. & Hedges, J. I. Sources and reactivities of amino acids in a coastal marine environment. *Limnol. Oceanogr.* **37**, 703–724 (1992).
17. Tanoue, E., Nishiyama, S., Kamo, M. & Tsugita, A. Bacterial membranes: possible source of dissolved protein in seawater. *Geochim. Cosmochim. Acta* **59**, 2643–2648 (1995).
18. Keil, R. G. & Kirchman, D. L. Dissolved combined amino acids in marine waters as determined by a vapor-phase hydrolysis method. *Mar. Chem.* **33**, 243–259 (1991).
19. van Heemst, J. D. H., Baas, M., de Leeuw, J. W. & Benner, R. in *Organic Geochemistry* (ed. Oygard, K.) 694–698 (Falch Høgtitrykk, Oslo, 1993).
20. Derenne, S., Largeau, C. & Taulelle, F. Occurrence of non-hydrolysable amides in the macromolecular constituent of *Scenedesmus quadricauda* cell wall as revealed by ^{15}N NMR: Origin of n-alkyl nitriles in pyrolysates of ultralaminae-containing kerogens. *Geochim. Cosmochim. Acta* **57**, 851–857 (1992).
21. De Leeuw, J. W. & Largeau, C. in *Organic Geochemistry* (eds Engle, M. & Macko, S. A.) 23–72 (Plenum, New York, 1993).
22. Rogers, H. J. Peptidoglycans: structure, function and variations. *Ann. New York Acad. Sci.* **235**, 29–51 (1974).
23. Aluwihare, L., Repeta, D. & Chen, R. A major biopolymeric component to dissolved organic carbon in seawater. *Nature* **387**, 166–169 (1997).
24. Bergamaschi, B. A. *The Marine Geochemistry of Carbohydrates: Application and Development of New Techniques of Analysis*, 1–184 (University of Washington, Seattle, 1995).
25. Patience, R. I. et al. The functionality of organic nitrogen in some recent sediments from the Peru upwelling region. *Org. Geochem.* **18**, 161–169 (1992).
26. Benzing-Purdie, L., Ripeester, J. A. & Preston, C. Elucidation of Nitrogen forms in melanoidins and humic acid by nitrogen-15 cross polarization magic angle-spinning nuclear magnetic resonance spectroscopy. *J. Agric. Food Chem.* **31**, 913–915 (1983).
27. Maillard, L. Formation de matieres humiques par action de polypeptides sur les sucres. *C. R. Acad. Sci.* **156**, 148–149 (1912).
28. Williams, P. M. & Druffel, E. R. M. Radiocarbon in dissolved organic matter in the central North Pacific Ocean. *Nature* **330**, 246–248 (1987).
29. Knicker, H. & Ludemann, H. N-15 and C-13 CPMAS and solution NMR studies of N-15 enriched plant material during 600 days of microbial degradation. *Org. Geochem.* **23**, 329–341 (1995).
30. Ripeester, J. A., Hawkins, R. E., MacPhee, J. A. & Nandi, B. N. On the interaction between pyridine and coal studies by CPMAS ^{15}N NMR. *Fuel* **65**, 740–742 (1986).

31. Benner, R. & Strom, M. A critical evaluation of the analytical blank associated with DOC measurements by high-temperature catalytic oxidation. *Mar. Chem.* **41**, 153–160 (1993).
 32. Hedges, J. I. & Stern, J. H. Carbon and nitrogen determinations of carbonate-containing solids. *Limnol. Oceanogr.* **29**, 657–663 (1983).

Acknowledgements. We thank the captain and the crew of the RV *John Vickers* for assistance with obtaining Equatorial Pacific samples, B. Black for help with sample processing, G. Cowie for advice in amino-acid analysis, and D. Bear for guidance and support.

Correspondence should be addressed to M.M. (e-mail: mdmccar@u.washington.edu).

Variability of the North Atlantic thermohaline circulation during the last interglacial period

Jess F. Adkins*§, Edward A. Boyle*, Lloyd Keigwin† & Elsa Cortijo‡

* Department of Earth Atmosphere and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

‡ Centre des Faibles Radioactivités, Laboratoire mixte CNRS-CEA 91198, Gif-sur-Yvette Cedex, France

Studies of natural climate variability are essential for evaluating its future evolution. Greenland ice cores suggest that the modern warm period (the Holocene) has been relatively stable for the past 9,000 years^{1,2}. Much less is known about other warm interglacial periods, which comprise less than 10% of the climate record during the past 2.5 million years^{3–7}. Here we present high-resolution ocean sediment records of surface and deep-water variables from the Bermuda Rise spanning the last interglacial period, about 118,000–127,000 years ago. In general, deep-water chemical changes are coincident with transitions in surface climate at this site. The records do not show any substantial fluctuations relative to the much higher variability observed during the preceding and subsequent cool climates. The relatively stable interglacial period begins and ends with abrupt changes in deep-water flow. We estimate, using ²³⁰Th measurements to constrain the chronology, that transitions occur in less than 400 years.

Early results from the GRIP ice core project suggested large climatic changes during the last interglacial^{2,8}, taken here as roughly equivalent to marine oxygen isotopic stage 5e and the Eemian interglacial on land⁹. Because of structural studies of the ice sheet's bedding¹⁰, comparison with the GISP2 record¹ and studies of the $\delta^{18}\text{O}$ of atmospheric oxygen¹¹, the largest of these apparent climatic instabilities now appear to be artefacts of ice disturbance near the bottom of the core. Nevertheless, the Greenland ice sheet data highlighted how little is known of the climate stability of warm climate periods such as stage 5e, inspiring detailed climate studies on land^{6,7} and in marine records^{3–5,12,13}. As part of the IMAGES coring program, we raised a 52.7-m-long core (MD95-2036, 33° 41.444'N, 57° 34.548'W, 4,462 m water depth) from the Bermuda Rise, a North Atlantic sediment drift deposit. This core has a continuous record of sedimentation from the Last Glacial Maximum back through the penultimate glaciation (Fig. 1). Visible light reflectance data from the core (a proxy for %CaCO₃) show that we have collected a record that includes all of the major interstadials from 5 to 21 that have been previously identified in Greenland ice cores. Extremely high deposition rates (10–200 cm kyr⁻¹) and penetration through the marine isotope stage 6 provide a unique archive of the ocean's behaviour. As has been previously reported¹⁴, this site is a sensitive indicator of basin scale deep-water chemical variability. Here we present results on the foraminiferal chemistry

and ²³⁰Th-based clay and carbonate accumulation rates in this core from the depth interval from 42 to 44 m, including marine stages 5e and 5d. To achieve the highest possible temporal resolution, the core was sampled at 1-cm intervals, taking care to avoid extraneous material from burrow structures. Although this sample spacing may seem very fine, it is important to note that bulk density data show that an original depth interval of ~2.4 cm has been compacted to 1 cm.

Several geochemical tracers are used here (Fig. 2). $\delta^{18}\text{O}$ measurements of the planktonic foraminiferan *Globigerinoides ruber* reflect a mixture of the global ice volume signature and local surface temperature and salinity conditions. Similarly, benthic $\delta^{18}\text{O}$ data reflect deep-water temperatures and the $\delta^{18}\text{O}$ value of deep water at this site (determined by global ice volume and deep water hydrography). For the global deep ocean, the $\delta^{18}\text{O}$ variation between glacial and interglacial extrema is usually considered to be about two-thirds ice volume and about one-third temperature¹⁵ (although one study argues for a larger role for temperature in the South Atlantic¹⁶). Cd/Ca ratios were obtained from the benthic foraminiferan *Nutallides umbonifera*. These data reflect the relative mixture of nutrient-depleted source waters of a northern origin (low Cd/Ca) with nutrient-enriched source waters of a southern origin (high Cd/Ca)^{14,17,18}. The foraminiferal dissolution index responds to the corrosiveness of bottom waters to CaCO₃. Because *Cibicides wuellerstorfi* abundance is extremely low in several zones, the benthic $\delta^{13}\text{C}$ record is incomplete. These $\delta^{13}\text{C}$ data are given in the Supplementary Information, but for deep-water chemical variability we depend instead on the more complete Cd/Ca data¹⁹. Spectrophotometric reflectance data from the raw sediment were obtained immediately on core splitting²⁰; these data are summarized with CIE (Commission Internationale de l'Éclairage) 'L' (lightness) and 'a' (red-to-yellow colour balance) indices. Lightness reflects the relative abundance of white CaCO₃ and (typically) darker non-carbonate aluminosilicate material, and the red-to-yellow colour balance reflects the abundance of haematite-rich sediment from the Maritime Provinces of Canada^{21,22}. Finally, clay and carbonate fluxes were calculated from %CaCO₃ measurements and from excess ²³⁰Th^o (where ^o stands for decay-corrected) determined, on every other sample, by inductively coupled plasma mass spectrometry (ICP-MS). This calculation is based on the precisely known production rate of ²³⁰Th in the water column by dissolved uranium (in flux units of d.p.m. cm⁻² kyr⁻¹) and the efficient removal of this ²³⁰Th to the underlying sediment. On a basin-wide scale, the concentration of ²³⁰Th in surface sediments (in units of d.p.m. g⁻¹) is inversely proportional to the total sediment input to the basin (in flux units of g cm⁻² kyr⁻¹). After accounting for decay since deposition at the surface, down-core accumulation rates of individual sediment components can then be calculated by dividing the ²³⁰Th^o expected by the concentration measured and multiplying by the component concentration^{23–25}.

An age model for the interval studied was developed from $\delta^{18}\text{O}$ stratigraphy and ²³⁰Th-constrained flux variations. Two time control points are taken from the Martinson *et al.*²⁶ orbitally tuned benthic isotope stack: the beginning of the 5/6 transition (4,380 cm = 131 kyr BP) and the 5d/5e transition (4,280 cm = 114 kyr BP). Previous work at this site established that the sediment composition is a balance between biogenic carbonate (and a small amount of atmospheric dust) falling out of the ocean surface and aluminosilicates from the continental margin transported into the interior of the basin by deep-sea currents²⁷. The sedimentation rate at this site is enhanced by lateral sediment focusing, where freshly deposited material is removed from some areas of the seafloor by deep-sea currents and then transported to this site. Low-resolution data from the Holocene²⁷ and our unpublished studies based on detailed ¹⁴C dates at this site for the past 3,000 yr indicate that the basin-wide clay flux and the local focusing factor are correlated, probably linked by seafloor circulation characteristics that control both margin erosion and lateral transport. The ²³⁰Th focusing factor

§ Also at MIT/WHOI Joint Program in Oceanography E34-209, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

(F) is the ratio of total excess $^{230}\text{Th}^\circ$ found at a site divided by the amount expected from production in the overlying water column. Normally F is calculated by integrated excess $^{230}\text{Th}^\circ$ measured between two well-dated sediment horizons and dividing by the amount produced during that time interval. Here, instead, we use $\%\text{CaCO}_3$ as a proxy for variations in F and use the excess $^{230}\text{Th}^\circ$ measurements to calculate ages in 2-cm intervals. This process allows us to adjust our age model for the changes in accumulation rate between stable-isotope age control points.

We assume that there is a linear relation between F and $\%\text{CaCO}_3$ during stage 5e (as seen in the Holocene and deglaciation data) and that the slope is as close to the Holocene value as it can be subject to the constraints below. The slope ($m = -6.7$) and intercept ($b = 56\%$) of this relation are chosen so that

$$\int_{t=114 \text{ kyr}}^{131 \text{ kyr}} F(t)dt = F_{\text{average}}(131 \text{ kyr} - 114 \text{ kyr})$$

In practice, we evaluate the integral at our sampling resolution of 2 cm and use our proxy generated F values to calculate the ages (Δt) between these depths. In this way, the age model is improved from an assumption of constant sedimentation between benthic isotope control points:

$$\sum \frac{^{230}\text{Th}^\circ_{\text{ex}} \rho \Delta z}{\beta(\text{water depth})} = F_{\text{average}} T = \sum (m\%\text{CaCO}_3 + b)\Delta t$$

$$\sum \Delta t = T$$

$$F \geq 1$$

Here the summation is done over the depth range for which there are $^{230}\text{Th}^\circ_{\text{ex}}$ measurements. T is the time interval, from benthic isotope control points, of this depth range; β is the water-depth-dependent ^{230}Th production rate constant; and ρ is the dry bulk density. The slope and intercept of the first equation are chosen as described above and such that F is always greater than one. This constraint of no winnowing (consistent with all measured F from this site) coupled with different Holocene and stage 5e $\%\text{CaCO}_3$ maxima causes the F - $\%\text{CaCO}_3$ relationship to be different for the two time periods (Holocene values: $b = 35\%$, $m = -0.5$). These differences are inconsequential in relation to the conclusions discussed below. If we assumed the extreme case of constant F between our two stable isotope age control points, the age of the terminal stage 5e event would increase by 2 kyr and the duration of this event would increase from 410 to 830 yr. Above the zone with ^{230}Th data, a constant accumulation rate of 11 cm kyr^{-1} is assumed, to be consistent with sedimentation rates immediately below and inferred ages of isotopic stages 5a and 5c (as determined from the optical reflectance stratigraphy).

At the site of MD95-2036, Cd data from the four oldest samples indicate that a significant component of low-nutrient North Atlantic Deep Water (NADW) was present at this site during early termination II (Fig. 2), although Cd is somewhat higher (lower $\%\text{NADW}$)

than seen during stage 5e or the Holocene. The largest signal in the benthic Cd data show that stage 5e is preceded by an abrupt cessation of NADW supply to this site lasting about 1,500 yr, coeval with the Glacial to Interglacial transition (Fig. 2). This may be the same event observed by Oppo *et al.*¹³ in a core shallower and farther to the north. A similar event has also been reported at this site for early deglaciation near the end of isotope stage 2 (refs 14, 19). All of these events may be coeval with 'Heinrich' ice rafting events. At the end of the termination II event, Cd values are low and similar to Holocene values,¹⁴ and benthic $\delta^{18}\text{O}$ attains values characteristic of full interglacial conditions. The bottom-water saturation state, as indicated by the fragmentation index, closely follows the Cd/Ca record, possibly with a slight time lag. This fragmentation record supports Cd/Ca evidence for NADW variability at this site. The clay flux peaks early on termination II, just preceding NADW disappearance from this site. This peak clay flux probably reflects higher sediment influx from Canada during glacial disintegration. The fine-grained aluminosilicates from this eroded material are first deposited on the Laurentian Fan,²² scoured by the western boundary current, and then laterally transported on to the Bermuda Rise.²⁸ The coincident enhanced reddish colour of the sediment—signifying hematite from the Canadian Maritimes—supports this interpretation. The data from the 5/6 transition show that the steepest change in $\delta^{18}\text{O}$ and the NADW retreat from this site are synchronous, but also show that clay flux precedes this transition and the deep-water change.

Planktonic and benthic $\delta^{18}\text{O}$ are well correlated throughout stages 5e and 5d; both attain full interglacial values at $\sim 129 \text{ kyr BP}$ and remain relatively constant for the next 8,000 years. Fine-scale structure in the planktonic record shows three zones of $\delta^{18}\text{O}$ variability: (1) values of about $-0.79 \pm 0.16\text{‰}$, with a slight decreasing trend during stage 5e; (2) more scattered values of $\sim -0.25 \pm 0.21\text{‰}$ during stage 5d; and (3) generally increasing values on the 5e to 5d transition. The climatic record at this site does not show a switch towards glacial conditions until there is a chemical change in the deep waters. Late within isotope stage 5e ($\sim 118 \text{ kyr BP}$), there is a rapid shift in oceanic conditions in the western North Atlantic that has not been previously documented. Within a 2-cm interval, $\%\text{CaCO}_3$ drops abruptly, and clay flux, Cd/Ca, percentage fragmentation and red/yellow colour balance all rise equally abruptly. These data indicate a coincident increase both in the proportion of southern source waters and the recirculation-derived clay supply. Immediately afterwards, benthic $\delta^{18}\text{O}$ begins a steady increase into marine isotope stage 5d. The rise in Cd/Ca, carbonate flux and red-to-yellow colour balance are short-lived, occurring over $< 8 \text{ cm}$ ($\sim 400 \text{ yr}$), whereas the clay flux decreases more gradually. Surface tracers (*G. ruber* $\delta^{18}\text{O}$ and CaCO_3 flux) are coincident with deep tracers (Cd/Ca and clay flux) but with a smaller amplitude. The rapid deep and surface hydrographic changes found in this core mark the end of the peak interglacial and the beginning of climate deterioration towards the semi-glacial

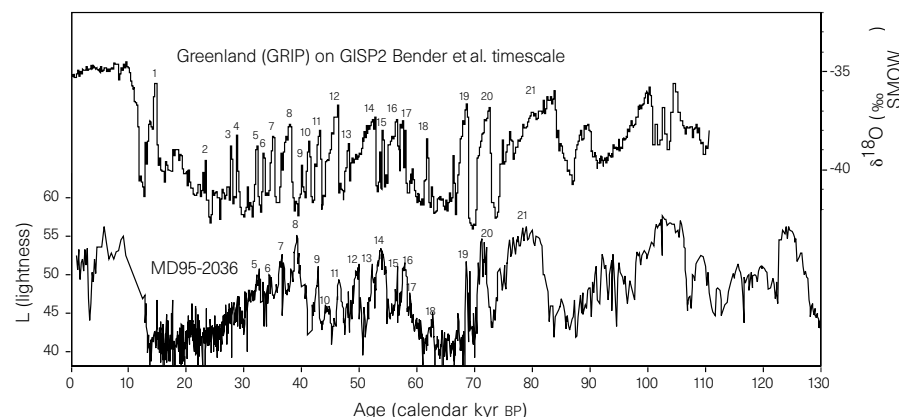


Figure 1 Bermuda Rise sediment grey scale compared with GRIP ice core $\delta^{18}\text{O}$. Lightness data were collected for MD95-2036 as described in the text. Ice core $\delta^{18}\text{O}$ data are from the GRIP ice core project.² The age model is obtained by correlating the GISP II $\delta^{18}\text{O}$ -age record with the GRIP $\delta^{18}\text{O}$ -depth record. Numbered intervals correspond to interstadial events identified in the ice core record. All events from 5 to 21 are identifiable in the marine cores, indicating a continuous record of climate up to the penultimate glaciation. SMOW, standard mean ocean water.

stage 5d. Before and immediately after this event, signalling the impending end of stage 5e, deep-water chemistry is similar to modern NADW. As climate deteriorates into stage 5d, the deep circulation becomes more variable ($Cd/Ca_{5d} = 0.099 \pm 0.021$ and $Cd/Ca_{5e} = 0.071 \pm 0.010$).

Data from the Bermuda Rise reveal several important features of the climate stability during stage 5e and the penultimate glacial transition. By augmenting the $\delta^{18}O$ stratigraphy with high-resolution ^{230}Th profiling, we can place our records on a more refined absolute age scale. During stage 5e, we find climate stability lasting $\sim 8,000$ yr in both surface and deep waters at the Bermuda Rise. Although there are significant events during this warm period, the

variance of our climate tracers during stage 5e is smaller than in the following cold stage and the preceding glacial transition. This stability is punctuated, at the end of the last interglacial plateau, by an abrupt event that begins the transition to a harsher climate. Over ~ 400 yr, Cd/Ca ratios, clay flux, percentage fragmentation and the red/yellow colour balance all show a synchronous switch to bottom waters that have a higher proportion of southern source component. It is this deep-water event that marks the descent into isotopic stage 5d. Our data indicate that climatic shifts are strongly related to deep-water reorganizations^{12,13} and that these changes can occur in a time of the order of a few hundred years or less. □

Received 24 March; accepted 8 September 1997.

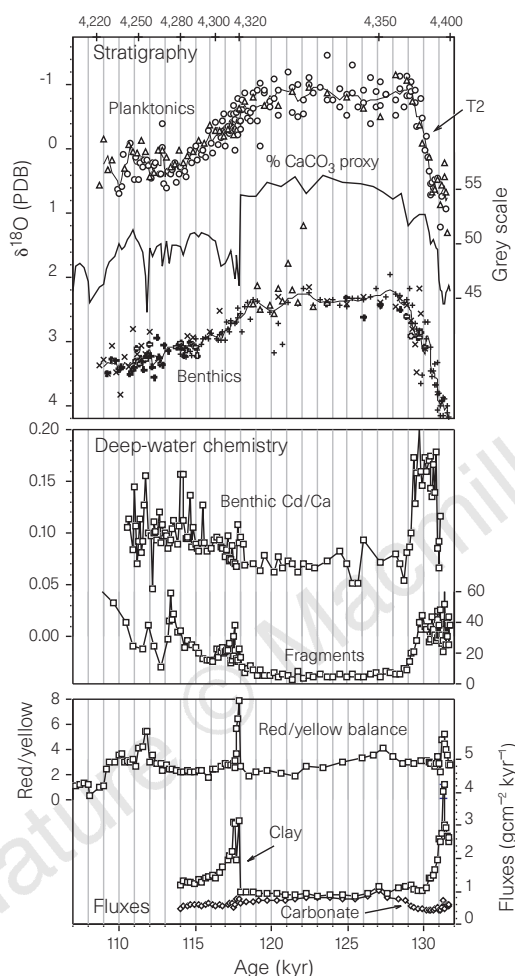


Figure 2 Plots of MD95-2036 data against age and depth. All isotopic data are relative to Pee Dee belemnite (PDB). T2 refers to the position of termination II, the penultimate deglaciation. The symbols in the planktonic stable isotope plot are VG $\delta^{18}O$ from *G. ruber* (circles) and Finnigan $\delta^{18}O$ from *G. ruber* (triangles). A three-point running average is drawn through the data. Benthic $\delta^{18}O$ data are from four sources: VG *Cibicides* (upside-down triangles), Finnigan *Cibicides* (plus signs), *Melonis* (squares) and *Oridorsalis* (circles). The *Melonis* and *Oridorsalis* data were adjusted by -0.35% and -0.3% respectively to be equivalent to the *Cibicides* data. These adjustments are consistent with offsets at depths where the species coexist with *Cibicides* and with previous core top studies.^{29,30} A *Cibicides* equivalent line is drawn through the data as a three-point running average. Cd/Ca ratios are from *N. umbonifera* and are plotted along with the foraminiferal fragmentation index. Carbonate and non-carbonate fluxes from excess ^{230}Th measurements are plotted with the reflectance red/yellow balance. The age model was calculated from benthic $\delta^{18}O$ and $^{230}Th_{ex}$ as described in the text. Note that although foraminiferal $MnCO_3$ levels in this deep section of the core are higher than normally considered acceptable ($\sim 200 \mu mol kg^{-1}$), the similarity of low-Mn Holocene and stage-5e Cd values suggests that these numbers are not negatively affected at this site.

- Groote, P. M., Stuiver, M., White, J. W. C., Johnsen, S. & Jouzel, J. Comparison of oxygen isotope records from GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
- GRIP Climate instability during the last interglacial period recorded in the GRIP ice core. *Nature* **364**, 203–207 (1993).
- Cortijo, E. et al. Eemian cooling in the Norwegian Sea and North Atlantic Ocean preceding continental ice growth. *Nature* **372**, 446–449 (1994).
- McManus, J. F. et al. High-resolution climate records from the North Atlantic during the last interglacial. *Nature* **371**, 326–329 (1994).
- Maslin, M., Sarnthein, M. & Knaack, J.-J. Subtropical Eastern Atlantic climate during the Eemian. *Naturwissenschaften* **83**, 122–126 (1996).
- Thouveney, N. et al. Climate variations in Europe over the past 140 kyr deduced from rock magnetism. *Nature* **371**, 503–506 (1994).
- Field, M. H., Huntley, B. & Muller, H. Eemian climate fluctuations observed in a European pollen record. *Nature* **371**, 779–783 (1994).
- Dansgaard, W. et al. Evidence of general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
- Shackleton, N. J. The last interglacial in the marine and terrestrial records. *Proc. R. Soc. Lond. B* **174**, 135–154 (1969).
- Alley, R. B. et al. Comparison of deep ice cores. *Nature* **373**, 393–394 (1995).
- Fuchs, A. & Leuenberger, M. C. $\delta^{18}O$ of atmospheric oxygen measured on the GRIP ice core document stratigraphic disturbances in the lowest 10% of the core. *Geophys. Res. Lett.* **23**, 1049–1052 (1996).
- Keigwin, L. D., Curry, W. B., Lehman, S. J. & Johnsen, S. The role of the deep ocean in North Atlantic climate change between 70 and 130 kyr ago. *Nature* **371**, 323–326 (1994).
- Oppo, D. W., Horowitz, M. & Lehman, S. Marine core evidence for a sharp decrease in deep water production during Termination II and a relatively stable MIS 5e (Eemian). *Paleoceanography* **12**, 51–64 (1997).
- Boyle, E. A. & Keigwin, L. D. North Atlantic thermohaline circulation during the last 20,000 years linked to high latitude surface temperature. *Nature* **330**, 35–40 (1987).
- Shackleton, N. J. & Opydyke, N. D. Oxygen isotope and paleomagnetic stratigraphy of equatorial Pacific core V28-238: oxygen isotope temperatures and ice volumes on 10^6 yr scale. *Quat. Res.* **3**, 39–55 (1973).
- Schrag, D. P., Hampt, G. & Murry, D. W. Pore fluid constraints on the temperature and oxygen isotopic composition of the Glacial ocean. *Science* **272**, 1930–1932 (1996).
- Hester, K. & Boyle, E. A. Water chemistry control of cadmium in recent benthic foraminifera. *Nature* **298**, 260–262 (1982).
- Boyle, E. A. & Keigwin, L. D. Deep circulation of the North Atlantic over the last 200,000 years: geochemical evidence. *Science* **218**, 784–787 (1982).
- Keigwin, L. D., Jones, G. A., Lehman, S. J. & Boyle, E. A. Deglacial meltwater discharge, North Atlantic deep circulation, and abrupt climate change. *J. Geophys. Res.* **96**, 16811–16826 (1991).
- Boyle, E. A. Characteristics of the deep ocean carbon system during the past 150,000 years: ECO_2 distributions, deep water flow patterns, and abrupt climate change. *Proc. Natl Acad. Sci. USA* **94**, 8300–8307 (1997).
- Balsam, W. L., Otto-Bliesner, B. L. & Deaton, B. C. Modern and last glacial maximum eolian sedimentation patterns in the Atlantic Ocean interpreted from sediment iron oxide content. *Paleoceanography* **10**, 493–507 (1995).
- Hollister, C. D. & McCave, I. N. Sedimentation under deep-sea storms. *Nature* **309**, 220–225 (1984).
- Bacon, M. P. & Rosholt, J. N. Accumulation rates of $Th-230$, $Pa-321$, and some transition metals on the Bermuda Rise. *Geochim. Cosmochim. Acta* **46**, 651–666 (1982).
- Bacon, M. P. Glacial to interglacial changes in carbonate and clay sedimentation in the Atlantic Ocean estimated from ^{230}Th measurements. *Isotope Geosci.* **2**, 97–111 (1984).
- Francois, R., Bacon, M. P. & Suman, D. O. ^{230}Th profiling in deep-sea sediments; high-resolution records of changes in flux and dissolution of carbonate in the equatorial Atlantic during the last 24,000 years. *Paleoceanography* **5**, 761–781 (1990).
- Martinson, D. G. et al. Age dating and the orbital theory of the ice ages: development of high-resolution 0 to 300,000-year chronostratigraphy. *Quat. Res.* **27**, 1–29 (1987).
- Suman, D. O. & Bacon, M. P. Variations in Holocene sedimentation in the North American basin determined from ^{230}Th measurements. *Deep-Sea Res.* **36**, 869–878 (1989).
- Laine, E. P. & Hollister, C. D. Geological effects of the Gulf Stream system on the northern Bermuda Rise. *Mar. Geol.* **39**, 277–300 (1981).
- Belanger, P. E., Curry, W. B. & Matthews, R. K. Core-top evaluation of benthic foraminiferal isotopic ratios for paleo-oceanographic interpretations. *Palaogeogr. Palaoclimatol. Palaeoecol.* **33**, 205–220 (1981).
- Graham, D. W., Corliss, B. H., Bender, M. L. & Keigwin, L. D. Carbon and oxygen isotopic equilibria of recent deep-sea benthic foraminifera. *Mar. Micropaleont.* **6**, 483–497 (1981).
- Bender, M. et al. Climate correlations between Greenland and Antarctica during the past 100,000 years. *Nature* **372**, 663–666 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial offices of Nature.

Acknowledgements. This research would not have been possible without the extraordinary efforts of the entire IMAGES 101 crew during the summer of 1995 on board the *Marion Dufresne*. Specifically we thank Yvon Balut, Laurent Labeyrie, Jean-Louis Turon and Franck Bassinot for their efforts on our behalf. We also thank M. M. Rutgers van der Loeff for helpful comments. J.F.A. was supported by a NASA Global Change Fellowship and a grant from the Tokyo Electric and Power Company. NSF provided the funds to collect this core and NOAA funded our work on high-resolution climate records.

Correspondence and requests for materials should be addressed to J.F.A. or E.A.B.

Displacement rates of normal faults

A. Nicol^{*†}, J. J. Walsh^{*}, J. Watterson^{*} & J. R. Underhill[†]

^{*} Fault Analysis Group, Department of Earth Sciences, University of Liverpool, Liverpool L69 3BX, UK

[†] Department of Geology and Geophysics, The University of Edinburgh, Grant Institute, West Mains Road, Edinburgh EH9 3JW, UK

Previous estimates of displacement rates on individual faults have been limited to neotectonic faults and averaged over time intervals of about 200 kyr or less^{1–5}. These estimates have been highly variable, which has led to a belief that longer-term displacement rates on individual faults are likely to be variable as well. Here we report estimates of long-term normal-fault displacement rates averaged over time intervals ranging from 1 to 40 Myr, and based on observed decreases in displacement of progressively younger horizons intersected by syn-sedimentary faults. We find that displacement rates are remarkably stable over these longer time periods, and within a given fault system the rates are strongly dependent on the relative size of the fault (as measured by cumulative vertical displacement). Taken together, these results indicate that faults become large relative to nearby faults by having higher displacement rates, even when small, rather than as a consequence of having been active for longer. Our analyses also show that high regional strain rates tend to be accommodated by high fault displacement rates rather than high fault densities.

We have estimated fault displacement rates of 86 syn-sedimentary normal faults in six regions, time-averaged over intervals of 1–40 Myr, which are significant proportions of the active lifetime of each fault. The data allow the effects on displacement rates of fault size and of basinal extension rate to be identified separately. The aim was to determine both the absolute fault displacement rates and the relative rates of displacement on faults within a fault system, and so address the basic question of whether large faults become larger than small faults because they moved faster, even when still small, or simply because they were active for longer. Fault size is expressed in terms of maximum cumulative throw (that is, vertical component of displacement), which is >75% of the total displacement.

The data are for normal faults from three offshore (Gulf Coast, North Sea and Timor Sea) and three onshore (Aegean, Basin & Range and Kenya Rift) extensional terrains (Fig. 1). The faults sampled are of tectonic origin except for the gravity-driven Gulf Coast faults. Within individual data sets the range of maximum displacements is generally at least one order of magnitude and the range over all data sets is 30 m to 6 km (Fig. 2). Thickening of a depositional unit across a syn-sedimentary fault, such as the Smith Bank fault shown in Fig. 1, is expressed by the stratigraphic growth index (hanging-wall thickness/footwall thickness) and ranges from ~1.16 to 2.5 for the offshore data sets. The gradients of throw upwards along the slip surfaces of these syn-sedimentary faults, 0.16–1.5, are higher than those on post-depositional faults of comparable size⁶ by a factor of at least 2. Biostratigraphic and radiometric dating of horizons shows that the relevant parts of the syn-faulting sequences were deposited over intervals of ~1–40 Myr.

High quality two-dimensional and three-dimensional seismic reflection data were used to analyse 57 faults from the offshore regions, in which hanging-wall sedimentation rates exceeded maximum fault displacement rates by factors of up to 30. Footwall erosion is insignificant and the growth indices are therefore direct reflections of the incremental fault throws during deposition of the syn-faulting sequences. Displacement backstripping^{7–10} was used to determine, for each fault, the difference in throw between the

youngest and oldest reliably dated horizons in the syn-faulting sequence. The greater precision of the Timor Sea and an Inner Moray Firth subset of the North Sea data allowed the definition of several growth stages of individual faults and the backstripping of several horizons.

Published displacement and age data were analysed for 29 faults in the Aegean¹¹, Basin and Range^{12–16} and Kenya Rift^{17,18} onshore regions. As the Kenya Rift faults are subject to footwall erosion, valid displacement rates can be calculated only for large displacement faults. Results for onshore faults are generally less robust than those for offshore faults.

Elapsed time plotted against maximum displacement is shown for 86 faults from the six regions on a log-log plot (Fig. 2). Rates of displacement vary with location and with fault size; the range of rates for individual faults is 0.004–1.0 mm yr⁻¹, which is similar to that for published rates averaged over much shorter intervals of 20–200 kyr, that is, 0.01–1.61 mm yr⁻¹ (ref. 5). Data distributed along a line parallel to the displacement axis would indicate a direct relation between fault size and displacement rate, and a distribution parallel to a line of slope 1 would correspond to a rate unrelated to size. The data sets (Fig. 2) show linear trends with slopes between 0 and 1, suggesting a correlation between displacement rate and fault size, but the data generally are too scattered for the relations between rates and fault sizes to be determined precisely. Precise relations between fault sizes and displacement rates were determined from the more refined data available for the Timor Sea and for the Inner

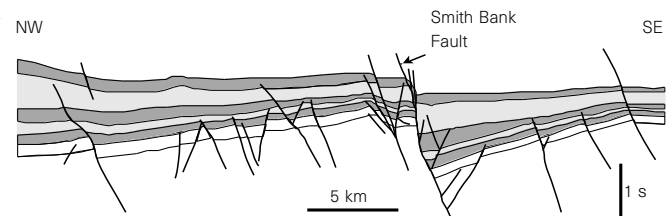


Figure 1 Cross section, interpreted from a two-dimensional seismic reflection line, of the Smith Bank Fault and associated Late Jurassic to Early Cretaceous growth strata (shaded), Inner Moray Firth, UK North Sea (after refs 20, 21). The growth sequence is the syn-rift package and has been subdivided into five discrete units (indicated by shading) that range in age from 144.5 to 133 Myr (refs 20, 21). The vertical axis is in two-way travel time and the vertical exaggeration is ~2.

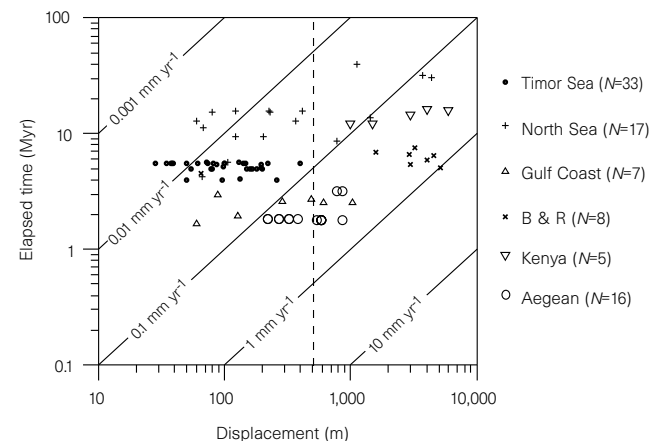


Figure 2 Log-log plot of elapsed time against maximum displacement for faults in three offshore (Timor Sea, North Sea and Gulf Coast) and three onshore (Basin and Range (B&R), Kenya Rift and Aegean) regions. For details see the text. Measurement errors of maximum throw for offshore and onshore data sets are estimated to be +1–10% and ±30%, respectively. Elapsed time data are estimated to be within ±30%. Oblique lines are contours of constant growth rate (mm yr⁻¹).

[†] Present address: Institute of Geological and Nuclear Sciences, PO Box 30 368, Lower Hutt, New Zealand.

Moray Firth subset of the North Sea data, where time resolutions are ~ 1 Myr (ref. 19) and ~ 3 Myr (refs 20, 21), respectively, and displacement rates can be calculated for several intervals within the syn-faulting sequences. The outstanding features of the plots of cumulative displacement against elapsed time (Fig. 3) are the relative scarcity of crossing growth profiles, the near-constant relative growth rates of all faults in each system and, perhaps less clearly, the near-linear growth profiles of individual faults. Even in a single fault system there is therefore no particular displacement rate for faults when they have attained a displacement of, say, 100 m; faster faults simply reach a given size sooner than slower faults.

The constant relative displacement rates on faults within a single fault system, on time scales greater than ~ 1 Myr, together with the near-constant absolute displacement rates on individual faults, are unexpected results. A possible explanation is that the syn-sedimentary growth of the faults followed an earlier period of basement faulting during which a wide size range of faults was established in terms of both displacement and dimensions, with subsequent syn-sedimentary growth by slip increments proportional in size to the dimensions of the pre-existing basement faults²². In this event the systematics evident from Fig. 3 would represent a special case, which might be acceptable for the Moray Firth fault system but not for the Timor Sea, where many of the faults are restricted to the syn-faulting sequence. In this more general case, initial size differences between faults must have been established during the nucleation and early growth of the fault system, within a period less than the ~ 1 Myr time resolution.

Whatever the origin of their initial size distributions, proportional as opposed to absolute size differences between faults are maintained during subsequent growth of the systems, requiring a positive feedback process by which initially larger faults maintain their relative size advantage. A mechanism for a competitive process of this type might be the release of elastic strain energy from larger rock volumes by faults that initially were either slightly larger or more isolated than competitor faults.

A representative displacement rate for each region (see the legend to Fig. 4) was obtained by estimating the rate of displacement on a nominal 500 m displacement fault, using the time value of the intersection between a double regression line for each data set (not shown) with the vertical line corresponding to 500 m displacement (Fig. 2). The characteristic 500 m fault displacement rate is plotted against the regional strain rate for each region in Fig. 4. Regional strain rates for the North Sea^{9,23}, the Aegean^{24,25}, the northern Basin and Range^{16,26–29} and for the Kenya Rift^{17,30} are derived from published sources.

The broad positive correlation between fault displacement rates and regional strain rates shows that differences between the regional strain rates are accommodated by near-proportional differences in fault displacement rates, rather than by different densities of active faults. If the differences were accommodated solely by fault displacement rates, the data in Fig. 4 would lie along a straight line with a slope of 1.0, whereas the slope of the best-fit line is < 1.0 . Given their uncertainties, the data are consistent with this simple interpretation but do not exclude the possibility that some other factor accommodates regional strain rate differences.

In the offshore data sets all faults were active at an early stage of the faulting deformation. The conclusion might be drawn that, as displacement rates on individual faults are almost constant, active fault populations will remain the same throughout the faulting event, as long as the regional strain rate is constant. However, as the maximum displacement on a fault increases, the dimensions also increase^{22,31}, so a fault with constant displacement rate accommodates a progressively higher proportion of the regional strain rate as its dimensions increase; that is, the rate of seismic moment on a growing fault increases even when the displacement rate is constant. A constant regional strain rate can be accommodated either by establishing fault lengths at early stages of growth of the system and

slowing lateral propagation by fault interaction³², or by decreasing the number of active faults with time, probably by the inactivation of smaller faults. A large initial number of faults, accommodating a distributed regional strain, followed by the progressive concentration of strain on to the larger faults, is characteristic of the North Sea data area, including the Inner Moray Firth (Fig. 1), and is also a feature of numerical rupture models^{33,34}. Many small faults (throw < 100 m) terminate upwards in the lower part of the North Sea syn-faulting sequence, with the youngest syn-faulting horizons offset only by the main intrabasin faults.

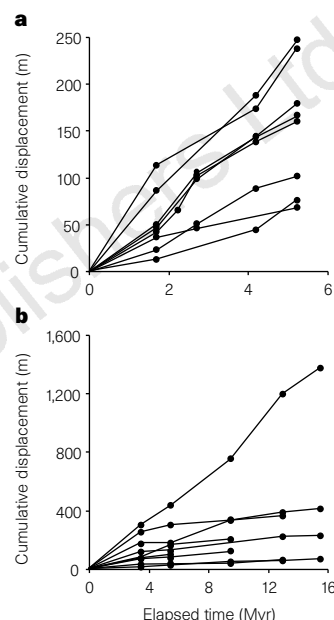


Figure 3 Cumulative displacement plotted against elapsed time for 17 faults from the Timor Sea ($N = 8$) (a), and the Inner Moray Firth ($N = 9$) (b). Axes represent the maximum fault throws and times elapsed since the onset of syn-faulting sedimentation. Curves were generated from cross sections recording the maximum displacement on each fault. Estimated error ranges for cumulative throws are ± 1 –10%. Elapsed times have precisions better than ~ 0.25 Myr for the Timor Sea and ~ 0.50 Myr for the Inner Moray Firth.

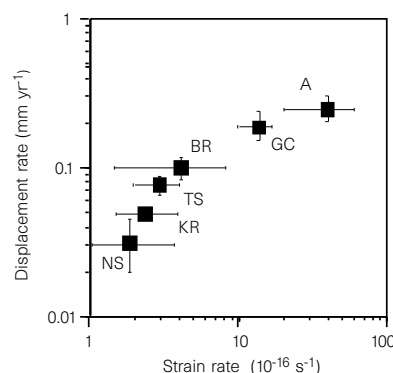


Figure 4 Log-log plots of displacement rates on nominal 500 m faults plotted against regional strain rate for Gulf Coast (GC), North Sea (NS), Timor Sea (TS), Aegean (A), northern Basin and Range (BR) and Kenya Rift (KR). Strain rate error bars show the ranges of estimates obtained from either regional cross sections (Gulf Coast, North Sea, Timor Sea and Kenya Rift), from seismic moment data (Aegean), or from both (Basin and Range). The strain rate shown for the Gulf Coast is a local rather than a regional value because the concept of a regional strain rate is not applicable to the Gulf Coast. Displacement rates of 0.188 (GC), 0.030 (NS), 0.074 (TS), 0.248 (A), 0.098 (BR) and 0.049 (KR) mm yr^{-1} were derived for each region with error bars defined by the envelope, parallel to the line of best fit, representing one standard deviation from the best-fit lines to the data in Fig. 2.

Palaeoseismic observations demonstrate temporal clustering of slip events on individual faults or fault segments over periods of up to ~200 kyr (see refs 2–4). This short-term clustering of slip events contrasts with the steady growth of individual faults, when averaged over longer intervals (>1 Myr), suggesting that punctuated displacement histories are short-term phenomena restricted to time scales of <1 Myr. Fault activity over longer time intervals and at regional scales might, however, be punctuated either temporally or spatially by external factors.

Our analysis shows that regions with higher strain rates have proportionally higher fault displacement rates on faults of all sizes, rather than proportionally greater numbers of faults. Within the Timor Sea and Moray Firth systems, displacement rates on individual faults are positively correlated with fault size, and larger faults are those that have grown faster. Consequently, the large faults were relatively large throughout the growth of each of these fault systems. The rates of displacements on individual faults are therefore dependent primarily on their relative sizes at an early stage of growth of the faults system and on the regional strain rate. Relatively constant fault growth rates require long-range correlation and interaction between all faults in a fault system, a characteristic of critical state systems and self-organized criticality. However, the physical basis for the emergence of these systematics is still uncertain. □

Received 20 February; accepted 10 September 1997.

- Schwartz, D. P. & Coppersmith, K. J. Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas Fault Zones. *J. Geophys. Res.* **89**, 5681–5698 (1984).
- Wallace, R. Patterns and timing of late Quaternary faulting in the Great Basin Province and relation to some regional tectonic features. *J. Geophys. Res.* **89**, 5763–5769 (1984).
- Wallace, R. E. Grouping and migration of surface faulting and variations in slip rates on faults in the Great Basin Province. *Bull. Seism. Soc. Am.* **77**, 868–876 (1987).
- Coppersmith, K. J. On spatial and temporal clustering of paleoseismic events. *Seismol. Res. Lett.* **59**, 299–230 (1989).
- McCalpin, J. P. Frequency distribution of geologically determined slip rates for normal faults in the western United States. *Bull. Seism. Soc. Am.* **85**, 1867–1872 (1995).
- Gillespie, P. A., Walsh, J. J. & Watterson, J. J. Limitations of dimension and displacement data from single faults and the consequences for data analysis and interpretation. *J. Struct. Geol.* **14**, 1157–1172 (1992).
- Petersen, K., Clausen, O. R. & Korstgård, J. A. Evolution of a salt-related listric growth fault near the D-1 well, block 5605, Danish North Sea: displacement history and salt kinematics. *J. Struct. Geol.* **14**, 565–577 (1992).
- Childs, C. et al. Kinematic analysis of faults in a physical model of growth faulting above a viscous salt analogue. *Tectonophysics* **228**, 313–329 (1993).
- Clausen, O. R. et al. in *Modelling the Earth for Oil Exploration* (Final Report of the CEC's Geoscience Program 1990–1993) (ed. Helbig, K.) 205–316 (Elsevier, 1994).
- Childs, C., Watterson, J. & Walsh, J. J. Fault overlap zones within developing normal fault systems. *J. Geol. Soc. Lond.* **152**, 535–549 (1995).
- Doutos, T. & Poulimenos, G. Geometry and kinematics of active faults and their seismotectonic significance in the western Corinth–Patras rift (Greece). *J. Struct. Geol.* **14**, 689–699 (1992).
- Zoback, M. L. in *Tectonic and Stratigraphic Studies in the Eastern Great Basin* (eds Miller, D. M., Todd, V. R. & Howard, K. A.) 3–28 (Geological Society of America, Memoir 157, 1983).
- Stein, R. S., King, G. C. P. & Rundle, J. B. The growth of geological structures by repeated earthquakes. 2. Field examples of continental dip–slip faults. *J. Geophys. Res.* **93**, 13319–13331 (1988).
- Anders, M. H., Geissman, J. W., Piety, L. A. & Sullivan, J. T. Parabolic distribution of circum-eastern Snake River Plain seismicity and latest Quaternary faulting: migratory pattern and association with the Yellowstone hotspot. *J. Geophys. Res.* **94**, 1589–1521 (1989).
- Pierce, K. L. & Morgan, L. A. in *Regional Geology of Eastern Idaho and Western Wyoming* (eds Link, P. K., Kuntz, M. A. & Platt, L. B.) 1–54 (Geological Society of America, Memoir 179, 1992).
- Anders, M. H., Spiegelman, M., Rodgers, D. W. & Hagstrum, J. T. The growth of fault-bounded tilt blocks. *Tectonics* **12**, 1451–1459 (1993).
- Chapman, G. R., Lippard, S. J. & Martyn, J. E. The stratigraphy and structure of the Kamasia Range, Kenya Rift Valley. *J. Geol. Soc. Lond.* **135**, 265–281 (1978).
- Jones, W. B. Listric growth faults in the Kenya Rift Valley. *J. Struct. Geol.* **10**, 661–672 (1988).
- Woods, E. P. Vulcan sub-basin fault styles—implications for hydrocarbon migration and entrapment. *Austr. Petrol. Explor. Ass. J.* **32**, 138–158 (1992).
- Underhill, J. R. Controls on Late Jurassic seismic sequences, Inner Moray Firth, UK North Sea: a critical test of a key segment of Exxon's original global cycle chart. *Basin Res.* **3**, 79–98 (1991).
- Underhill, J. R. Implications of Mesozoic–Recent basin development in the western Inner Moray Firth, UK. *Mar. Petr. Geol.* **8**, 359–369 (1991).
- Walsh, J. J. & Watterson, J. Analysis of the relationship between the displacements and dimensions of faults. *J. Struct. Geol.* **10**, 239–247 (1988).
- Barr, D. 3-D palinspastic restoration of normal faults in the Inner Moray Firth: implications for extensional basin development. *Earth Planet. Sci. Lett.* **75**, 191–203 (1985).
- Jackson, J. & McKenzie, D. Rates of active deformation in the Aegean Sea and surrounding regions. *Basin Res.* **1**, 121–128 (1988).
- Le Pichon, X. & Angelier, J. The Hellenic arc and trench system: a key to the neotectonic evolution of the eastern Mediterranean area. *Tectonophysics* **60**, 1–42 (1979).
- Wright, L. Late Cenozoic fault patterns and stress fields in the Great Basin and westward displacement of the Sierra Nevada Block. *Geology* **4**, 489–494 (1976).
- Thompson, G. A. & Burke, D. B. Regional geophysics of the Basin and Range province. *Annu. Rev. Earth Sci.* **2**, 213–238 (1974).
- Eaton, G. P., Wahl, R. R., Probst, H. J., Mabey, D. R. & Kleinkopf, M. D. in *Cenozoic Tectonics and Regional Geophysics of the Western Cordillera* (eds Smith, R. B. & Eaton, G. P.) 51–91 (Geological Society of America, Memoir 152, 1978).

- Eddington, P. K., Smith, R. B. & Renggli, C. in *Continental Extensional Tectonics* (eds Coward, M. P., Dewey, J. F. & Hancock, P. L.) 371–392 (Special Publication 28, Geological Society, London, 1987).
- Baker, B. H. & Wohlenberg, J. Structure and evolution of the Kenya Rift Valley. *Nature* **229**, 538–542 (1971).
- Cowie, P. A. & Scholz, C. H. Physical explanation for displacement—length relationship for faults using a post-yield fracture mechanics model. *J. Struct. Geol.* **14**, 1133–1148 (1992).
- Nicol, A., Watterson, J., Walsh, J. J. & Childs, C. The shapes, major axis orientations and displacement patterns of fault surfaces. *J. Struct. Geol.* **18**, 235–248 (1996).
- Cowie, P. A., Vanneste, C. & Sornette, D. Statistical physics model for the spatiotemporal evolution of faults. *J. Geophys. Res.* **98**, 21809–21821 (1993).
- Cowie, P. A., Sornette, D. & Vanneste, C. Multifractal scaling properties of a growing fault population. *Geophys. J. Int.* **122**, 457–469 (1995).

Acknowledgements. We thank members of the Liverpool Fault Analysis Group for useful discussions; staff of BHP Petroleum, Exxon and Statoil for providing access to seismic and other data; and Geco-Prakla for releasing seismic data from the Inner Moray Firth. We also thank Conrad Childs and Graham Yielding for comments on early versions of this article, and Mark Anders and Brian Wernicke for their reviews. This research was partly funded by the DTI/NERC Hydrocarbon Reservoirs LINK Programme.

Correspondence and requests for materials should be addressed to J.J.W. (e-mail: fault@fag.esc.liv.ac.uk).

Pb, U and Th diffusion in natural zircon

James K. W. Lee^{*†}, Ian S. Williams^{*} & David J. Ellis[‡]

^{*} Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia

[‡] Department of Geology, The Australian National University, Canberra, ACT 0200, Australia

Zircon (ZrSiO₄) is one of the most widely used minerals for determining the age, origin and thermal history of rocks by U–Th–Pb geochronology. But the parameters describing the solid-state (volume) diffusion rates of these elements in natural zircon, which are themselves important for establishing the limits on the applicability of zircon for geochronological studies, have remained poorly quantified¹. This is because of the measurement difficulties associated with the low (p.p.m.) concentrations and low diffusion rates of these elements in natural zircon, and the chemical and physical heterogeneity present in most crystals. Here we present direct measurements of the uranium, thorium and lead loss from a thermally treated gem-quality natural zircon and show that lead diffuses much faster than uranium or thorium. We find that the U–Th–Pb isotopic system in natural zircon will typically have a closure temperature greater than 900 °C, which explains why zircon is apparently such a robust geochronometer and is capable of remaining isotopically closed through extended periods of high-grade metamorphism and partial melting of the host rock.

We have used a variety of analytical techniques (scanning electron microscopy (SEM), transmission electron microscopy (TEM), atomic force microscopy (AFM), and secondary ion mass spectrometry (SIMS)) to overcome the above difficulties and to measure directly the diffusion coefficients of Pb, U and Th in a well-characterized natural zircon crystal, the Research School of Earth Sciences (RSES) zircon standard SL13. SL13 is a low-U (238 p.p.m.), low-Th (21 p.p.m.) gem-quality zircon megacryst from Sri Lanka originally selected for its relative uniformity in U concentration and ²⁰⁶Pb/²³⁸U on a 20-μm scale. Multiple isotope dilution analyses² yield a mean ²⁰⁶Pb/²³⁸U ratio of 0.0928 (²⁰⁶Pb = 19 p.p.m.), corresponding to an age of 572.1 Myr. TEM examination of the crystal showed that, despite minor radiation damage (α-dose ≈ 4.76 × 10¹⁴ decays per mg), the zircon exhibited excellent crystallinity (Fig. 1).

In preparation for the diffusion experiments, randomly oriented fragments of SL13 about 1 mm in diameter were polished with 0.05-μm alumina to produce one flat face and annealed in air at 800 °C for 24 hours³. For each experiment, two fragments were placed in a

[†] Present address: Department of Geological Sciences, Queen's University, Kingston, Ontario K7L 3N6, Canada.

platinum capsule with spectroscopically pure NaCl (4 ± 2 p.p.b. U, 35 ± 2 p.p.b. Th, and 485 ± 10 p.p.b. Pb, as determined by inductively coupled plasma mass spectrometry) and finely crushed pure synthetic zircon⁴ (U, Th, Pb <10 p.p.b.). The NaCl, when molten, provided a large chemical potential gradient for trace-element diffusion and thus acted as an ionic sink for diffusing cations. The synthetic zircon served as a buffer to minimize dissolution. These capsules were sealed *in vacuo* (pressure $P < 10^{-5}$ torr) and then heated at temperatures of 900, 1,000 and 1,100 °C respectively for 28 to 104 days (Table 1). Subsequent SEM examination of the SL13 fragments showed that the polished crystal surfaces remained intact—they had neither reacted nor partly dissolved.

The thermal treatment produced near-surface Pb, U and Th concentration gradients beneath the polished faces consistent with thermally induced diffusion. These gradients were measured by depth profiling with the SHRIMP II ion microprobe using a very weak primary ion beam (0.4–0.6 nA, O₂⁺) which was broadly focused (40-μm diameter) to achieve the maximum secondary ion yield at a minimum penetration rate. The ion yields were calibrated against untreated, polished and annealed SL13 fragments analysed under identical conditions; the useful yield (the ratio of Pb ions detected to sample Pb atoms sputtered) was ~1.2%. The pits produced by the ion probe analyses were imaged and measured by atomic force microscopy (AFM), which confirmed that the penetration depth over most of the analysed area was very uniform, and that the sputtering rates (typically 0.045–0.06 nm s⁻¹) were con-

stant. The time-dependent ion-probe analyses were thereby calibrated against penetration depth into the crystals. A typical depth profile at 1,100 °C, and its linearization by inversion through the error function, are illustrated in Fig. 1. Diffusion coefficients were calculated by least-squares fitting of the depth-profiling data to the solution of the diffusion equation for a semi-infinite medium with a uniform initial concentration and a surface concentration of zero⁵.

The results of the diffusion experiments are given in Table 1. Diffusion coefficients for Pb were obtained at all three temperatures; diffusion coefficients for U and Th could only be estimated from the data obtained at 1,100 °C. The analyses suggest that at 1,100 °C, Pb diffusion is faster than U or Th diffusion by about 4 orders of magnitude. Diffusion of U seems to be slightly faster than that of Th, which is consistent with diffusivity model predictions^{6,7} based on the ionic radii of the cations (U⁴⁺ = 1.00 Å, Th⁴⁺ = 1.05 Å)⁸, and a diffusion study using synthetic zircon⁹, but the limited data are not conclusive. Predicting the mobility of Pb is more problematic, first because the valence of Pb, which is not known in zircon, strongly controls its ionic radius (Pb²⁺ = 1.29 Å, whereas Pb⁴⁺ = 0.94 Å; ref. 8); and second because a Pb atom, being the product of a series of energetic α-decays, almost certainly will not be located or bound in a U or Th lattice site.

There are two key tests by which to determine whether the Pb concentration gradients can be described by a volume diffusion process. First, the diffusion coefficient (D) should be independent of the time (t) for which each sample was heated. Second, D as a

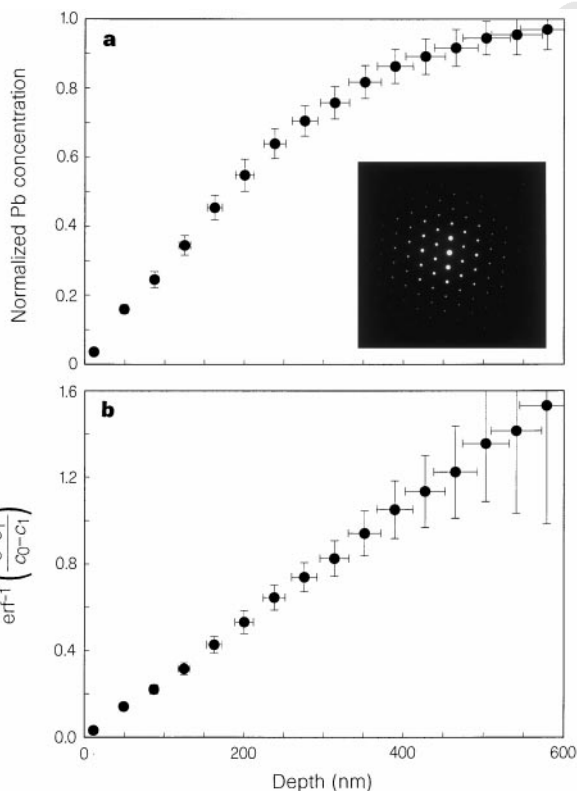


Figure 1 a, Typical Pb diffusion profile after heating at 1,100 °C. For diffusion out of a semi-infinite medium, the normalized concentration $(c - c_1)/(c_0 - c_1)$ is related to the depth x into the crystal by $(c - c_1)/(c_0 - c_1) = \text{erf}(x/2\sqrt{Dt})$, where c_0 is the initial concentration in the crystal, c_1 is the concentration at the crystal surface, c is the concentration at any depth, and $\text{erf}(x)$ is the value of the error function at x . Inset: TEM diffraction pattern of natural SL13 (untreated). Well-defined spots separated by spacings characteristic of the zircon unit-cell indicate excellent crystallinity. **b**, Linearization of **a** by inversion through the error function. Error bars for the ordinate increase significantly as the normalized concentration approaches 1 because of the exponential nature of the error function.

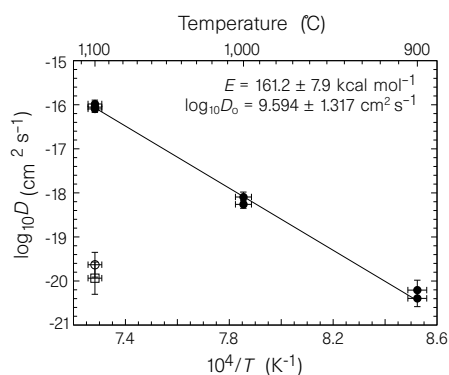


Figure 2 Arrhenius diagram of Pb, U and Th diffusion coefficients. Pb, black circles; U, open circle; Th, open square. Because the temperature dependence of D is defined by an Arrhenius relationship, a plot of $\log_{10} D$ against $1/T$ should yield a straight line. The line drawn through the Pb data represents a weighted least-squares linear fit with an activation energy E of 161 kcal mol⁻¹ and a pre-exponential coefficient D_0 of 3.9×10^9 cm² s⁻¹.

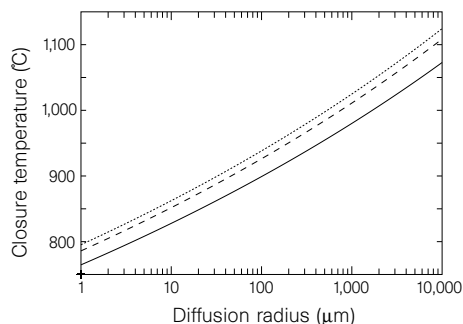


Figure 3 Zircon Pb closure temperature plotted against effective radius of diffusion. Curves are shown for cooling rates of 1 (solid line), 5 (dashed line) and 10 (dotted line) °C Myr⁻¹, using the fundamental diffusion parameters determined in this study.

Table 1 Diffusion data for U, Th and Pb in natural zircon

Sample	<i>T</i> (°C)	Time (×10 ⁹ s)	<i>D</i> _U (cm ² s ⁻¹)	<i>D</i> _{Th} (cm ² s ⁻¹)	<i>D</i> _{Pb} (cm ² s ⁻¹)	log ₁₀ <i>D</i>
Z2388-4	900	8.952			4.03 × 10 ⁻²¹	-20.395 ± 0.187
Z2477-3	900	6.051			6.21 × 10 ⁻²¹	-20.207 ± 0.225
Z2462-1	1,000	5.516			5.48 × 10 ⁻¹⁹	-18.261 ± 0.090
Z2517-2	1,000	3.117			8.05 × 10 ⁻¹⁹	-18.094 ± 0.112
Z2518-1	1,100	4.449			8.04 × 10 ⁻¹⁷	-16.095 ± 0.079
Z2527-3	1,100	2.867			1.04 × 10 ⁻¹⁶	-15.983 ± 0.088
Z2527-4	1,100	2.421			8.84 × 10 ⁻¹⁷	-16.054 ± 0.085
Z2518-1	1,100	4.449	2.34 × 10 ⁻²⁰	1.16 × 10 ⁻²⁰		-19.631 ± 0.281
Z2518-1	1,100	4.449				-19.936 ± 0.366

function of temperature (*T*) should be described by the Arrhenius relation,

$$D = D_0 e^{-E/RT}$$

where *E* is the activation energy, *D*₀ is the pre-exponential coefficient and *R* is the gas constant. Both tests are passed; experiments at constant *T* (900, 1,000, and 1,100 °C) lasting different times did yield similar values for *D* (Table 1), and the diffusion data from Table 1 display excellent Arrhenius behaviour (Fig. 2). Weighted least-squares linear regression of the data¹⁰ yields an activation energy *E* of 161 ± 8 kcal mol⁻¹ and a pre-exponential coefficient *D*₀ of 3.9^{+77.6}_{-3.7} × 10⁹ cm² s⁻¹. The possible dependence of the Pb diffusion rate on crystallographic orientation was not tested directly, but the consistent Arrhenius behaviour of *D*, derived from randomly oriented crystal fragments, suggests that Pb diffusional anisotropy is not significant.

The values of *D* for U and Th at 1,100 °C are so low (~10⁻²⁰ cm² s⁻¹) that those elements are unlikely to diffuse under most geological conditions, so the closure temperature of the Pb–U–Th isotopic system (the temperature of the system at the time represented by its apparent age¹¹) in zircon is determined by the diffusion rate of Pb. If one assumes a spherical diffusion geometry and that the zircon grain radius approximates the effective diffusion radius, then the closure temperatures calculated from the values of *D*₀ and *E* above for a zircon 200 µm in diameter which has cooled at rates of 1, 5 and 10 °C Myr⁻¹ are 899 ± 7, 926 ± 6 and 938 ± 5 °C, respectively. The relationships between the Pb closure temperature and the effective radius of diffusion for different cooling rates are shown in Fig. 3.

Zircon growth occurs in a variety of igneous, metamorphic and sedimentary environments over a wide range of temperatures, from diagenesis to magma genesis. Lead closure temperatures in excess of 900 °C imply that zircon Pb–U and Pb–Th isotopic ages are unlikely to be reset by thermal effects alone, except under the most extreme conditions of granulite-grade metamorphism or partial melting. Moreover, the closure temperature is above the temperature of many granite magmas (~750–850 °C), which explains the widespread preservation of inherited zircon within granites and high-grade metamorphic rocks. This indicates that the processes that govern the mobility of radiogenic Pb, and hence most isotopic discordance in zircon, are those that separately or in combination modify the zircon crystal lattice. Examples are radiation damage^{12,13}, fracturing due to differential lattice expansion¹⁴ and pressure release¹⁵, recrystallization^{16,17}, self-annealing, chemical reaction¹⁸ and leaching or alteration by fluids^{19–21}. Strong evidence for this is the common survival, within most discordant zircon populations, of some concordant zircon which can be dated after appropriate sample preparation (such as air abrasion or partial dissolution), or by SIMS techniques. This argument does not apply, of course, to the apparent discordance that results from the analysis of heterogeneous zircon populations or crystals of mixed age.

The rates of U and Th diffusion reported here are greater than those reported for U and Th in synthetic zircon⁹. An important experimental difference, however, is that we have measured profiles

of U, Th and Pb that have diffused out of natural zircon at natural concentrations, not profiles of elements diffusing into a pure synthetic crystal. The radiation damage caused by α-decay in natural crystals produces extended defects which may enhance diffusion rates in zircon. Therefore, the closure temperatures determined here should be considered maximum values for radiation-damaged zircon and minimum values for zircon annealed at modest temperatures (200–700 °C) over geological timescales. Lead closure temperatures much higher than those calculated here would be required to explain the negligible loss of radiogenic Pb from kimberlite zircon megacrysts postulated to have resided in the subcontinental mantle at temperatures in excess of 1,100 °C for ~2.55 Gyr (ref. 22). The diffusion parameters above yield a mean Pb migration distance of ~0.5 cm at 1,100 °C after 45 Myr, implying that even large (1 cm diameter) mantle zircons would lose most of their radiogenic Pb after this duration of heating. This suggests that if old mantle zircons formed within the diamond stability field in the upper mantle, soon afterwards they were transported to higher (cooler) levels in the lower crust. □

Received 21 April; accepted 21 September 1997.

- Lee, J. K. W. in *Defects and Processes in the Solid State* (eds Boland, J. N. & Fitz Gerald, J. D.) 423–446 (Elsevier, New York, 1993).
- Roddick, J. C. & van Breemen, O. in *Radiogenic Age and Isotopic Studies: Report 8*, Geol. Surv. Can. Curr. Res. 1994-F, 1–9 (1994).
- Reddy, K. P. & Cooper, A. R. Oxygen diffusion in sapphire. *J. Am. Ceram. Soc.* **65**, 634–638 (1982).
- Sinha, A. K., Wayne, D. M. & Essex, R. Flux growth of pure and doped zircons. *J. Cryst. Growth* **125**, 431–439 (1992).
- Crank, J. *The Mathematics of Diffusion* (Clarendon, Oxford, 1975).
- Dowty, E. Crystal-chemical factors affecting the mobility of ions in minerals. *Am. Mineral.* **65**, 174–182 (1980).
- Fortier, S. M. & Giletti, B. J. An empirical model for predicting diffusion coefficients in silicate minerals. *Science* **245**, 1481–1484 (1989).
- Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Cryst.* **A32**, 751–767 (1976).
- Cherniak, D. J., Hanchar, J. M. & Watson, E. B. Diffusion of tetravalent cations in zircon. *Contrib. Mineral. Petrol.* **127**, 383–390 (1997).
- York, D. Least squares fitting of a straight line with correlated errors. *Earth Planet. Sci. Lett.* **2**, 479–482 (1969).
- Dodson, M. H. Closure temperature in cooling geochronological and petrological systems. *Contrib. Mineral. Petrol.* **40**, 259–274 (1973).
- Silver, L. T. & Deutsch, S. Uranium–lead isotopic variations in zircons: a case study. *J. Geol.* **71**, 721–758 (1963).
- Silver, L. T. in *Nuclear Geophysics* (ed. Hurley, P. M.) Natl Acad. Sci.–Nat. Res. Council Publ. 1075, 34–39 (1963).
- Lee, J. K. W. & Tromp, J. Self-induced fracture generation in zircon. *J. Geophys. Res.* **100**, 17753–17770 (1995).
- Goldich, S. S. & Mudrey, M. G. in *Recent Contributions to Geochemistry and Analytical Chemistry* (ed. Tugarinov, A. I.) 466–470 (Wiley, New York, 1972).
- Black, L. P., Williams, I. S. & Compston, W. Four zircon ages from one rock: the history of a 3930 Ma-old granulite from Mount Sones, Enderby Land, Antarctica. *Contrib. Mineral. Petrol.* **94**, 427–437 (1986).
- Pidgeon, R. T. Recrystallisation of oscillatory zoned zircon: some geochronological and petrological implications. *Contrib. Mineral. Petrol.* **110**, 463–472 (1992).
- Watson, E. B. & Harrison, T. M. Zircon saturation revisited: temperature and composition effects in a variety of crustal magma types. *Earth Planet. Sci. Lett.* **64**, 295–304 (1983).
- Pidgeon, R. T., O'Neil, J. R. & Silver, L. T. Uranium and lead isotopic stability in a metamict zircon under experimental hydrothermal conditions. *Science* **154**, 1538–1540 (1966).
- Krogh, T. E. & Davis, G. L. Alteration in zircons with discordant U–Pb ages. *Ann. Rep. Dir. Geophys. Lab. Carnegie Inst. Washington Ybk* **73**, 560–567 (1973).
- Sinha, A. K., Wayne, D. M. & Hewitt, D. A. The hydrothermal stability of zircon: preliminary experimental and isotopic studies. *Geochim. Cosmochim. Acta* **56**, 3551–3560 (1992).
- Kinny, P. D., Compston, W., Bristow, J. W. & Williams, I. S. in *Kimberlites and Related Rocks*, vol. 2 (eds Ross, J. et al.) *Geol. Soc. Austral. Spec. Pub.* **14**, 833–842 (Blackwell, Melbourne, 1989).

Acknowledgements. We thank W. Compston and D. H. Green (Director, RSES) for supporting this project. We thank the following RSES technical personnel for assistance and advice: R. Willison for electron-beam welding; and R. Waterford, K. Massey, R. Willis and B. Hiberson for sample preparation.

Numerous people provided guidance and assistance with equipment: J. Fitz Gerald (TEM), T. Senden and A. Hyde (AFM), R. Headey (SEM), S. Eggins (ICP-MS) and D. Tilley (XRD). We also thank E. B. Watson for advice and discussions about experimental procedures and Pb diffusion, H. O'Neill and S. Chakraborty for discussions about thermodynamics and diffusion, and A. Sinha for providing us with synthetic zircon crystals. This work was supported by the Research School of Earth Sciences and an Australian Research Council grant to D. Ellis for the experimental work carried out in the Geology Department.

Correspondence should be addressed to J.K.W.L. (e-mail: lee@geol.queensu.ca).

Biodiversity regulates ecosystem predictability

Jill McGrady-Steed, Patricia M. Harris & Peter J. Morin

Department of Ecology, Evolution & Natural Resources, PO Box 231, Cook College, Rutgers University, New Brunswick, New Jersey 08903-0231, USA

Links between biodiversity and ecosystem function provide compelling reasons for conserving maximal numbers of species in ecosystems^{1–6}. Here we describe a previously unrecognized effect of biodiversity on ecosystem predictability, where predictability is inversely related to temporal and spatial variation in ecosystem properties. By manipulating biodiversity in aquatic microbial communities, we show that one process, ecosystem respiration, becomes more predictable as biodiversity increases. Analysis of similar patterns extracted from other studies^{2,3,6} indicates that biodiversity also enhances predictability in terrestrial ecosystems. Biodiversity can also affect average levels of ecosystem performance, but the extent to which different species make unique or redundant contributions to ecosystem processes remains controversial^{3,7–10}. Nonlinear effects of biodiversity on the decomposition of particulate organic matter and resistance of communities to invasion indicate that different species have redundant functions in our system. The consequences of biodiversity are also not restricted to early successional situations as described in previous studies^{1–4,6}, because strong effects persist even after ecosystems develop for periods corresponding to 40–80 generations of dominant organisms.

An unbiased experimental test of biodiversity's influence on ecosystem processes should create a biodiversity gradient by randomly assigning species to communities, to avoid confounding biodiversity with species composition. Plant communities are randomly assembled with relative ease^{4,6}, but plants are only one of many trophic levels in complex natural food webs, and interactions among trophic levels undoubtedly affect ecosystem patterns. Random assembly of multilevel food webs poses formidable practical problems because natural food webs are nonrandom collections of species¹¹, and because randomly assembled food webs often fail to persist¹². Instead of randomly assembling entire food webs, we used aquatic microbial communities to create a biodiversity gradient such as might result from species loss across all trophic levels of an initially diverse community^{2,3}. Although this approach risks confounding species richness with species composition, we provide additional experimental and statistical controls to separate these factors. We used several different species combinations to create alternative communities of identical species richness within the lower half of our biodiversity gradient (Table 1). We also used multiple regression analysis to assess whether species richness independently affected ecosystem attributes after statistically controlling for variation in species abundances^{1,5}. Differences in diversity rapidly stabilized and persisted for our study's 6-week duration (Fig. 1).

We used changes in the concentration of carbon dioxide within each microcosm as a convenient measure of ecosystem respiration. CO₂ flux is known to vary with species richness in other systems^{2,3} and indicates whether ecosystems are sources or sinks for this greenhouse gas. CO₂ production (measured as µl produced in

18 h) increased as species richness increased (Fig. 2a; Pearson's $r = 0.39$, $P = 0.0002$, $n = 85$), unlike the trend seen in terrestrial systems^{2,3}, where increased species richness resulted in a greater uptake of CO₂. Multiple regression analysis confirmed the independent influence of species richness on CO₂ flux in addition to effects of specific taxa (Table 2). The triangular scatter of observations (Fig. 2a) suggested that variation among replicates of identical species richness declined as species richness increased.

We further investigated relations between species richness and ecosystem variability using data consisting of observation made weekly for six weeks of CO₂ flux and species richness obtained for each microcosm. We grouped observations with the same species richness and calculated the standard deviation of CO₂ flux within each species richness group as a measure of ecosystem variability. This analysis combined temporal variation within replicates together with spatial variation among replicates to generate an overall measure of ecosystem variation (Fig. 2b). Ecosystem variation decreased as diversity increased (Spearman rank correlation between the standard deviation of CO₂ flux and realized species

Table 1 Species combinations used to create a biodiversity gradient

Trophic position	Organism	Initial species richness							
		0	3	5	10	15	20	25	31
Producers	<i>Ankistrodesmus</i> 1	(a)	c		a, c	a	a	a	a
	<i>Chlamydomonas</i>	(a)	a, b	a–c	a–d	a	a	a	a
	<i>Diatom</i> sp.			c, d	b, d				
	<i>Euglena</i>					a	a	a	a
	<i>Netrium</i>					a	a		a
	(<i>Phacus</i>)				a			a	a
	(<i>Peridinium</i>)				a		a	a	a
	<i>Scenedesmus</i>		d	a, d	b–d	a		a	a
	<i>Staurastrum</i>			b	b–d		a	a	a
	<i>Brachionus</i>		c		c	a			a
Herbivores	<i>Frontonia</i>		d		a, d		a	a	a
	<i>Hypostome</i> sp.		b	c	b		a	a	a
	<i>Stentor</i> 1			b	b		a	a	a
	<i>Stylonychia</i>			a, d	c		a	a	a
	<i>Aspidisca</i>							a	a
	(<i>Amoeba</i>)						a		a
	(<i>Coleps</i>)						a		a
Bacterivores	<i>Colpidium</i>			c	a, c	a	a	a	a
	<i>Colpoda</i> sm.							a	a
	<i>Colpoda</i> lg.							a	a
	<i>Halteria</i>		a		d	a	a	a	a
	<i>Gastrotrich</i> sp.				a, d	a			a
	Microflagellates	(a)	a–d	a–d	a–d	a	a	a	a
	<i>Monostyla</i>			a	b		a	a	a
	<i>P. bursaria</i>				b		a	a	a
	<i>Paramecium</i> 2			d	a, d	a		a	a
	<i>Rotaria</i>			b	c	a		a	a
Predators	<i>Spirostomum</i>				d	a		a	a
	<i>Heliozoa</i> sp.				b			a	a
	(<i>Oxytricha</i>)				a		a	a	a
	<i>Stentor</i> 2				c	a	a	a	a
	(<i>Urostyla</i>)					a	a		a

Letters a, b, c and d indicate the species composition of up to four different communities within each level of biodiversity. Taxa in parentheses failed to become established. Letters in parentheses in the 0 species treatment indicate organisms that contaminated microcosms originally containing only bacteria. Three other contaminants occurred sporadically in other treatments (2 amoebae and *Ankistrodesmus* 2).

richness; $r_s = -0.8286$, $P = 0.0001$, $n = 21$), indicating that ecosystem performance became increasingly predictable as species richness increased. Highly variable results at the lowest values of species richness were not the only cause of this pattern. Variability still correlated negatively with realized species richness ($r_s = -0.8006$, $P = 0.0005$, $n = 14$) within a subset of communities containing 8 or more species. Similar patterns exist in other studies, but their implications for ecosystem predictability have not been emphasized. We obtained measures of ecosystem variability (standard errors of the means) from two published studies of terrestrial ecosystems and correlated those measures with species richness. We found that ecosystem variability decreased as biodiversity increased for several processes: nutrient uptake in natural and planted grasslands⁶ ($r_s = -0.8567$, $P = 0.0001$; $r_s = -0.9543$, $P = 0.0008$), community respiration³ ($r_s = -0.8567$, $P = 0.0001$) and production³ ($r_s = -1.0$, $P = 0.0001$). We conclude that several properties of experimental ecosystems become more predictable as biodiversity increases. Recent models predict similar patterns in other systems where competing species are randomly assigned to communities to

create differences in biodiversity¹³. A subset of our data in which each initial diversity treatment corresponded to a unique species composition (series 'a' in Table 1) still shows an inverse relation between ecosystem variation and biodiversity (correlation between s.d. of CO₂ flux and species richness; $r_s = -0.6617$, $P = 0.0015$, $n = 20$). This pattern cannot be attributed to greater variation in species composition among low-diversity communities, as would result from random assignment of species to create each level of diversity. A similar pattern obtained in another study^{2,3} cannot be attributed to random community assembly either.

Another ecosystem process, decomposition of particulate organic matter, increased nonlinearly with species richness to attain maximal values in communities containing eight or more species (ANOVA $P = 0.0001$; Fig. 3a). This convex nonlinear pattern is consistent with functional redundancy among species in communities of moderate-to-high diversity. Multiple regression analysis using backward elimination of independent variables shows that decomposition depends on species richness as well as on the abundances of several bacterivores and predators (Table 2). Enhanced decomposition in species-rich communities was probably

Table 2 Summary of multiple regression analysis

CUMCO2	Parameter	P	PCTLOST	Parameter	P	INVADERS	Parameter	P
Intercept	-483.57	0.0105	Intercept	18.26	0.0002	Intercept	-0.0008	0.9800
Species richness	67.14	0.0017	Richness	2.07	0.0158	<i>Chlamydomonas</i>	0.0420	0.0001
<i>Ankistrodesmus</i> 1	-52.78	0.0436	<i>Monostyla</i>	11.58	0.0001	<i>Halteria</i>	-0.0356	0.0169
<i>Halteria</i>	288.15	0.0001	<i>Heliozoa</i>	-7.71	0.0319	<i>Stylonychia</i>	0.645	0.0025
<i>Hypostome</i>	456.28	0.0001	<i>Paramecium</i>	14.32	0.0001	<i>Scenedesmus</i>	0.0246	0.0031
<i>Ankistrodesmus</i> 2	-83.43	0.0378	<i>Stentor</i> 2	13.52	0.0028	<i>Staurastrum</i>	-0.0246	0.0035
<i>Stentor</i> 1	-1,062.21	0.0001				<i>Colpidium</i>	-0.0568	0.0057
<i>Euglena</i>	175.70	0.0411				<i>Paramecium</i>	-0.0444	0.0009
<i>Netrium</i>	-953.71	0.0111				<i>Stentor</i> 2	-0.0828	0.0086
<i>Aspidisca</i>	-503.71	0.0498						
$R^2 = 0.52$			$R^2 = 0.70$			$R^2 = 0.51$		

Multiple regressions (all significant at $P = 0.0001$) relate cumulative CO₂ flux (CUMCO2), decomposition (PCTLOST) and invasibility (INVADERS) to species richness and logarithmically transformed abundances of all species during week 6.

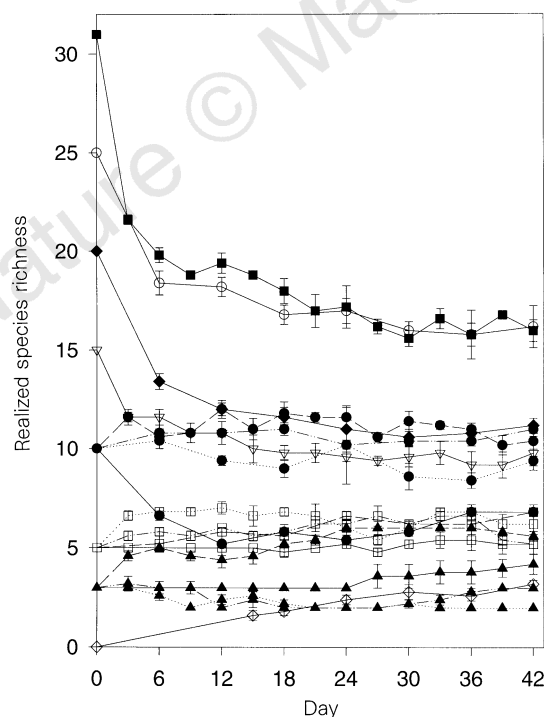


Figure 1 Mean realized species richness (± 1 s.e.) over time within biodiversity treatments. Species richness quickly stabilized and differences among treatments persisted. Symbols identify treatments that initially contained 0 (bacteria control; open diamonds), 3 (solid triangles), 5 (open squares), 10 (solid circles), 15 (open triangles), 20 (closed diamonds), 25 (open circles) or 31 (solid squares) eukaryotic species. Within an initial diversity level, different line types designate treatments containing different sets of species.

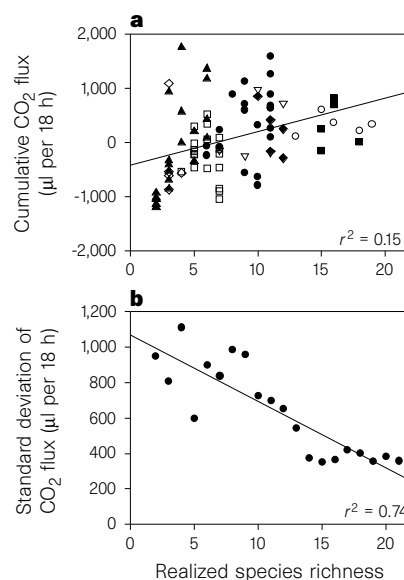


Figure 2 a, Relation between cumulative CO₂ flux and realized species richness after 6 weeks. Points are values of cumulative CO₂ flux in individual replicates. Different symbols identify different initial levels of biodiversity, as in Fig. 1. Communities with different species compositions but similar initial values of species richness are plotted using the same symbol. **b**, Relation between total variation in CO₂ flux (s.d. within groups of observations with the same species richness) and realized species richness over all 6 weeks. Variation described includes variation among replicates to provide an overall estimate of ecosystem predictability.

an indirect consequence of interactions between protists and the bacteria that cause decomposition. Protists did not directly consume the substrate used to measure decomposition. Enhanced bacterial turnover in communities of greater protist diversity could explain this pattern, especially if consumption by protists enhanced bacterial productivity. A positive correlation (Pearson's $r = 0.46$, $P = 0.0001$, $n = 85$) between CO_2 production (a measure of whole-community respiration that includes bacterial respiration) and decomposition during week 6 is consistent with this hypothesis.

Theory suggests that species will invade communities with increasing difficulty as biodiversity increases^{14–16}. We tested each community's resistance to invasion by adding small founding populations (ten cells) of *Euplotes* sp. to each microcosm after the sixth week of community development. This species is a facultative heterotroph containing endosymbiotic algae, and should consequently invade well even where heterotrophic resources are scarce. We also knew that *Euplotes* was not consumed by any of the predators remaining in the microcosms after six weeks of development, eliminating predation as a cause of failed invasions. We measured *Euplotes* population densities two weeks after their introduction, to permit at least 10–15 generations of invader population growth. *Euplotes* successfully invaded only communities containing fewer than six species (ANOVA $P = 0.0001$; Fig. 3b), but even some low biodiversity communities resisted invasion. Multiple regression analysis shows that species richness failed to account for invasibility after abundances of organisms on several trophic levels enter the regression (Table 2). The fact that communities with substantially less than the maximum observed biodiversity successfully resisted invasion is consistent with redundant effects of different species on invasibility. This pattern, although it is consistent with predictions of a positive relation between species richness and resistance to invasion^{14–19}, was primarily attributable to particular species whose abundances were correlated with species richness.

Undesirable changes in ecosystem processes provide one of the more compelling reasons for conserving biodiversity^{2–4,7–9,20}. This and other studies^{1–6} show that the greatest deterioration in ecosystem processes occurs as diversity declines from moderate (~8 species) to very low values. Average values of ecosystem processes generally level off between intermediate and high values of biodiversity and

remain unaffected by further increases in diversity. It is uncertain whether natural systems containing more species and experiencing greater environmental variation will exhibit similar patterns.

Knowledge of ecosystem predictability facilitates effective ecosystem management by providing insights into potential temporal variation within systems, as well as possible spatial variation among comparable systems. Because some processes continue to decrease in variability after average values reach an asymptote⁶, maintenance of high biodiversity is desirable even where species have largely redundant functional roles. The enhanced predictability of more complex systems, together with the functional deterioration that accompanies extensive loss of biodiversity, provide important reasons for conserving species richness in relatively intact ecosystems and for restoring diversity in degraded systems. □

Methods

Community assembly and measurement. We added bacteria, protists and small metazoans to aquatic microcosms to create 8 different levels of species richness (initially containing 0, 3, 5, 10, 15, 20, 25 or 31 eukaryotic species plus an average of 24 bacterial taxa; Table 1). We varied diversity by adding different species combinations to 100 ml of medium in 250-ml screw-capped Erlenmeyer flasks. Temporal staggering of species introductions by different trophic levels ensured that prey became abundant before predators were added. Realized species richness and densities of each species were estimated for each microcosm by identifying and counting eukaryotes from subsamples of known volume at regular intervals. Departures of realized species richness from initial values result from minor contamination (see treatment Oa in Table 1) as well as the loss of 6 other taxa. The various species combinations were run in two series (one labelled a, the other labelled b–d in Table 1) separated by several months, using otherwise identical culture conditions.

Decomposition. Change in the dry mass of preweighted sterile wheat seeds enclosed in nylon mesh bags placed in each microcosm measured decomposition. Seeds were extracted, dried and weighed after 6 weeks.

Invasion. Microcosms were subsampled every third day to follow invasion dynamics, after adding 10 *Euplotes* to each microcosm in week 6. The final abundance of *Euplotes* after 14 d provides an estimate of invasibility.

Respiration. A multichannel closed-circuit microrespirometer (Columbus Scientific Instruments) measured the cumulative production or consumption of carbon dioxide in each microcosm every week during an 18-h period under constant illumination.

Data analysis. Backward elimination multiple regression analyses related cumulative CO_2 flux (CUMCO₂), decomposition (PCTLOST), and invasibility (INVADERS) to species richness (RICHNESS) and logarithmically transformed abundances of all species during week 6. Only intercepts and variables with significant ($P < 0.05$) contributions to the model were retained. Retention of species richness in the regression model indicates an independent contribution of species richness in addition to the effects accounted for by variation in the densities of each species.

Received 3 March; accepted 4 August 1997.

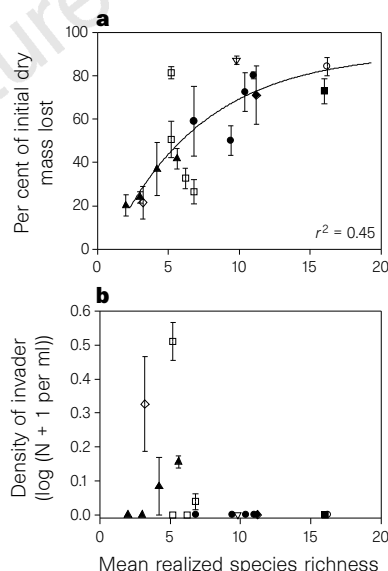


Figure 3 **a**, Decomposition (mean per cent loss dry g $\pm$ 1 s.e.) after 6 weeks versus average realized species richness. The best-fit equation for an exponential increase to a maximum is: decomposition = $98.29(1 - e^{-0.16(\text{species richness})}) - 9.32$. **b**, Invasibility (mean $\pm$ 1 s.e.), given by the logarithm of invader (*Euplotes*) density 2 weeks after introduction, versus mean realized species richness on the day of introduction. Symbols are as in Fig. 1.

1. Tilman, D. & Downing, J. A. Biodiversity and stability in grasslands. *Nature* **367**, 363–365 (1994).
2. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
3. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Empirical evidence that declining species diversity may alter the performance of terrestrial ecosystems. *Phil. Trans. R. Soc. Lond.* **347**, 249–262 (1995).
4. Naeem, S., Hakansson, K., Lawton, J. H., Crawley, M. J. & Thompson, L. J. Biodiversity and plant productivity in a model assemblage of plant species. *Oikos* **76**, 259–264 (1996).
5. Tilman, D. Biodiversity: population versus ecosystem stability. *Ecology* **77**, 350–363 (1996).
6. Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
7. Schulze, E. D. & Mooney, H. A. (eds) *Biodiversity and Ecosystem Function* (Springer, New York, 1993).
8. Johnson, K. H., Vogt, K. A., Clark, H. J., Schmitz, O. J. & Vogt, D. J. Biodiversity and the productivity and stability of ecosystems. *Trends Ecol. Evol.* **11**, 372–377 (1996).
9. Ehrlich, P. R. & Ehrlich, A. H. *Extinction. The Causes and Consequences of the Disappearance of Species* (Random House, New York, 1981).
10. Lawton, J. H. & Brown, V. K. in *Biodiversity and Ecosystem Function* (eds Schulze, E. D. & Mooney, H. A.) 255–270 (Springer, New York, 1993).
11. Pimm, S. L. Properties of food webs. *Ecology* **61**, 219–225 (1980).
12. Morin, P. J. & Lawler, S. P. Food web architecture and population dynamics: theory and empirical evidence. *Annu. Rev. Ecol. Syst.* **26**, 505–529 (1995).
13. Tilman, D., Lehman, C. L. & Thompson, K. T. Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
14. Pimm, S. L. in *Biological Invasions: A Global Perspective* (ed Drake, J. A. et al.) 351–367 (Wiley, New York, 1989).

15. Pimm, S. L. *The Balance of Nature?* (Univ. Chicago Press, Chicago, 1991).
16. Tilman, D. Community invasibility, recruitment limitation, and grassland biodiversity. *Ecology* **78**, 81–92 (1997).
17. Case, T. J. Invasion resistance arises in strongly interacting species-rich model competition communities. *Proc. Natl Acad. Sci. USA* **87**, 9610–9614 (1990).
18. Elton, C. S. *The Ecology of Invasions by Plants and Animals* (Chapman and Hall, London, 1958).
19. Vitousek, P. M. Biological invasions and ecosystem processes: towards an integration of population biology and ecosystem studies. *Oikos* **57**, 7–13 (1990).
20. Kareiva, P. M., Kingsolver, J. G. & Huey, R. B. (eds) *Biotic Interactions and Global Change* (Sinauer, Sunderland, Massachusetts, 1993).

Acknowledgements. We thank the NSF for supporting our research, and J. H. Lawton, D. Tilman, S. Naem, S. Lawler, M. Holyoak, P. Steed and T. Casey for their comments.

Correspondence and requests for materials should be addressed to P.J.M. (e-mail: pjmorin@rci.rutgers.edu).

Mice lacking bombesin receptor subtype-3 develop metabolic defects and obesity

Hiroko Ohki-Hamazaki[†], Kei Watase^{*}, Kazutoshi Yamamoto[‡], Hiroo Ogura[§], Mariko Yamano^{||}, Kazuyuki Yamada^{*}, Hiroshi Maeno^{*}, Junko Imaki[†], Sakae Kikuyama[‡], Etsuko Wada^{*} & Keiji Wada^{*}

^{*} Department of Degenerative Neurological Diseases, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo 187, Japan

[†] Department of Neurochemistry, Tokyo Institute of Psychiatry, Setagaya, Tokyo 156, Japan

[‡] Department of Biology, School of Education, Waseda University, Shinjuku, Tokyo 169-50, Japan

[§] Tsukuba Research Laboratories, Eisai Co. Ltd., Tsukuba, Ibaraki 300-26, Japan

^{||} Osaka Prefectural College of Health Sciences, Habikino, Osaka 583, Japan

[¶] Department of Anatomy, Nippon Medical School, Bunkyo, Tokyo 113, Japan

Mammalian bombesin-like peptides are widely distributed in the central nervous system as well as in the gastrointestinal tract, where they modulate smooth-muscle contraction, exocrine and endocrine processes, metabolism and behaviour¹. They bind to G-protein-coupled receptors on the cell surface to elicit their effects. Bombesin-like peptide receptors cloned so far include, gastrin-releasing peptide receptor (GRP-R)^{2,3}, neuromedin B receptor

(NMB-R)^{4,5}, and bombesin receptor subtype-3 (BRS-3)^{6,7}. However, despite the molecular characterization of BRS-3, determination of its function has been difficult as a result of its low affinity for bombesin and its lack of an identified natural ligand. We have generated BRS-3-deficient mice in an attempt to determine the *in vivo* function of the receptor. Mice lacking functional BRS-3 developed a mild obesity, associated with hypertension and impairment of glucose metabolism. They also exhibited reduced metabolic rate, increased feeding efficiency and subsequent hyperphagia. Our data suggest that BRS-3 is required for the regulation of endocrine processes and metabolism responsible for energy balance and adiposity. BRS-3-deficient mice provide a useful new model for the investigation of human obesity and associated diseases.

The gene that encodes BRS-3 is located on chromosome X and consists of three exons⁸. We generated male mice hemizygous for a deletion of the *BRS-3* gene by homologous recombination in embryonic stem cells (Fig. 1a, b). Wild-type and mutant mice were born at almost the same ratio. The disruption of the *BRS-3* gene in mutant mice was confirmed by reverse transcription-polymerase chain reaction (RT-PCR; Fig. 1c).

BRS-3-deficient mice were viable and fertile. They did not show any remarkable difference from wild-type mice until about 16 weeks of age, after which there was a distinct size difference between wild-type and mutant mice. When we compared the body weights of the F₂ generation, the weight gain of BRS-3-deficient mice was significantly greater than that of the wild-type mice (Fig. 2a). In all generations obtained so far by backcrossing to mice of the strain C57BL/6J (B6), mutant mice have differed significantly from wild type in body weight but not in length. To determine the tissue responsible for this weight gain, we weighed the adipose tissue in these mice. At 13 weeks of age, the adipose tissue mass, including epididymal, inguinal and retroperitoneal fatpads, was similar in wild-type and mutant mice (Table 1). However, at 27–34 weeks, when the weight difference was clear, the increased amount of white adipose tissue mass in BRS-3-deficient mice became apparent.

Heart rate and systolic and diastolic blood pressure were measured in these mice. Increases of 13% and 27% in systolic and diastolic blood pressure, respectively, were observed in 40-week-old BRS-3-deficient mice compared with age-matched wild-type mice

Table 1 Adipose tissue mass in BRS-3-deficient and control mice

	Body weight (g)	Epididymal fat-pad weight (% of body weight)	Inguinal fat-pad weight (% of body weight)	Retroperitoneal fat-pad weight (% of body weight)
Males (13 weeks)				
Control (n = 5)	25.4 ± 0.4	1.3 ± 0.1	1.5 ± 0.1	0.4 ± 0.1
Mutant (n = 5)	25.2 ± 1.0	1.4 ± 0.2	1.6 ± 0.2	0.4 ± 0.0
	P = 0.88	P = 0.58	P = 0.48	P = 0.88
% of control	99%	108%	110%	100%
Males (27–34 weeks)				
Control (n = 8)	36.0 ± 1.7	2.4 ± 0.3	2.6 ± 0.4	0.7 ± 0.1
Mutant (n = 7)	46.2 ± 1.8	3.5 ± 0.2	4.7 ± 0.4	1.1 ± 0.1
	P < 0.002	P < 0.04	P < 0.002	P < 0.02
% of control	128%	146%	181%	157%

After measurement of body weight, the mice were killed and white fat pads from the indicated sites were dissected and weighed. All values are means ± s.e.m. Statistics used a two-tailed unpaired Student's *t*-test.

Table 2 Heart rate and blood pressure in male BRS-3-deficient and control mice

	Body weight (g)	Heart rate (beats per min)	Systolic blood pressure (mm Hg)	Diastolic blood pressure (mm Hg)
Mice at 14–16 weeks				
Control (n = 6)	26.1 ± 1.2	682.2 ± 27.8	109.8 ± 3.8	65.3 ± 6.0
Mutant (n = 6)	28.1 ± 1.3	698.7 ± 14.5	111.0 ± 3.3	73.8 ± 4.1
Mice at 40 weeks				
Control (n = 6)	30.6 ± 2.2	709.2 ± 10.7	115.8 ± 2.6	75.2 ± 3.7
Mutant (n = 6)	44.8 ± 1.8	699.0 ± 19.1	131.2 ± 2.8	95.7 ± 2.1
	P < 0.0006		P < 0.004	P < 0.0008

Values are means ± s.e.m. Statistics used a two-tailed unpaired Student's *t*-test.

(Table 2). Various biochemical parameters were then determined in the blood of these mice (Table 3). The mice were fasted for 24 h, after which the blood glucose, cholesterol and triacyl glycerol levels were similar to those of the control mice, but the insulin level was elevated significantly in BRS-3-deficient mice at 23 weeks. We then examined metabolic changes that had an effect on the metabolism of glucose and fat (Table 4). Blood contents of thyroid hormone and glucagon were not perturbed, but secretion of growth hormone was diminished by 46%. Both insulin and growth hormone affect glucose transport, lipogenesis and lipolysis, and are important for the regulation of body lipid storage and use. Hyperinsulinaemia and low secretion of growth hormone are observed in other genetically obese mice and hypothalamic obese rats⁹; our

results suggest that glucose and lipid metabolism are similarly perturbed in BRS-3-deficient mice.

To assess the response to glucose administration, we performed oral glucose tolerance tests in these mice (Fig. 3). BRS-3-deficient mice at 26 weeks of age had higher levels of blood glucose and insulin than wild-type mice 15 min after oral glucose administration. Impairment of glucose metabolism was further confirmed by an insulin tolerance test (Fig. 4). After injection of insulin, the blood glucose level at 15–60 min in BRS-3-deficient mice of 38 weeks of age remained significantly higher than in wild-type mice, the inference being that BRS-3-deficient mice of this age could not clear glucose as efficiently as normal controls.

Because all the murine models of obesity exhibit hyperphagia, we

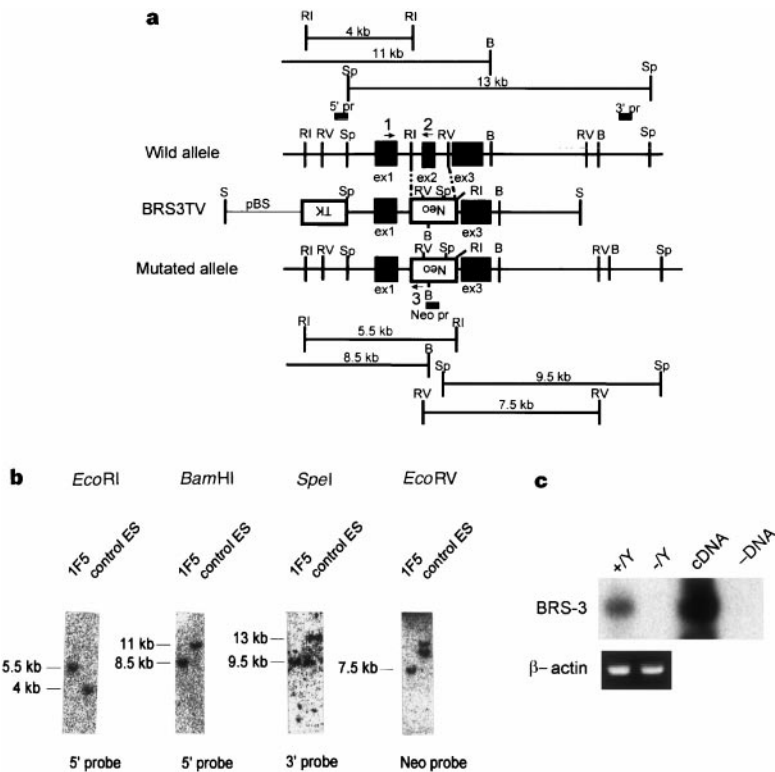


Figure 1 Targeted disruption of the *BRS-3* gene. **a**, Structure of the wild-type allele, gene-targeting vector (BRS3TV), and mutated allele. BRS-3 is encoded by 3 exons. Predicted lengths of restriction fragments for diagnosis are shown. Numbers 1, 2 and 3 indicate the position of PCR primers used in genotyping. RI, *EcoRI*; B, *Bam*HI; Sp, *SpeI*; RV, *EcoRV*; S, *SalI*. **b**, Southern blot analysis of DNA prepared from homologous recombinant ES (1F5) and control ES clone. **c**, Lack of expression of the *BRS-3* gene in mutant mice was confirmed by RT-PCR⁸. cDNA, positive control reaction using *BRS-3* cDNA; -DNA, negative control reaction without DNA.

Table 3 Blood analysis on fasted BRS-3-deficient and control mice

Males	Body weight (g)	Glucose (mg dl ⁻¹)	Insulin (ng ml ⁻¹)	Cholesterol (mg dl ⁻¹)	Triacyl glycerol (mg dl ⁻¹)
5 weeks					
Control (n = 8)	19.7 ± 0.5	102.7 ± 4.4	0.72 ± 0.21	n.d.	n.d.
Mutant (n = 8)	20.9 ± 0.4	104.0 ± 4.7	0.59 ± 0.17	n.d.	n.d.
8–10 weeks					
Control (n = 12)	24.5 ± 0.7	101.0 ± 3.8	0.57 ± 0.07	n.d.	n.d.
Mutant (n = 12)	25.5 ± 0.4	105.3 ± 5.4	0.79 ± 0.09	n.d.	n.d.
	P = 0.24		P < 0.06		
23 weeks					
Control (n = 6)	33.6 ± 2.1	136.4 ± 13.5	0.58 ± 0.21	99.2 ± 6.9	80.9 ± 21.3
Mutant (n = 6)	40.8 ± 0.7	146.1 ± 7.1	1.00 ± 0.23	111.0 ± 11.4	94.7 ± 10.3
	P < 0.02		P < 0.04	P = 0.38	P = 0.58

After a 24-h fast, blood was collected from the tail. Values are means ± s.e.m. Statistical analysis used a two-tailed unpaired Student's *t*-test. n.d., not determined.

Table 4 Serum metabolites and hormone levels in BRS-3-deficient and control mice

	Body weight (g)	Glucose (mg dl ⁻¹)	Insulin (ng ml ⁻¹)	Glucagon (pg ml ⁻¹)	Growth hormone (ng ml ⁻¹)	Free T ₃ (pg ml ⁻¹)	Free T ₄ (ng dl ⁻¹)
Control	31.0 ± 1.1 (n = 12)	297.5 ± 17.8 (n = 11)	1.84 ± 0.36 (n = 10)	69.3 ± 2.9 (n = 3)	8.9 ± 1.7 (n = 10)	2.7 ± 0.4 (n = 3)	0.74 ± 0.04 (n = 4)
Mutant	35.3 ± 1.1 (n = 12)	303.1 ± 13.8 (n = 11)	3.05 ± 0.56 (n = 11)	62.7 ± 5.4 (n = 3)	4.8 ± 0.6 (n = 11)	3.0 ± 0.2 (n = 3)	0.77 ± 0.01 (n = 4)
	P < 0.02		P = 0.10		P < 0.04		

Blood from free-fed animals at 19–20 weeks was collected by cardiac puncture under anaesthesia. Values are means ± s.e.m. Statistics used a two-tailed unpaired Student's *t*-test.

measured the food and water consumption of the BRS-3-deficient mice (Fig. 2b). At 10–12 weeks of age, when mutant mice started to gain excess weight, food intake was similar to that of control mice, and water intake was reduced. However, at 20–24 weeks, when obesity progressed, BRS-3-deficient mice ingested about 9% more calories than control mice. We monitored food consumption and change in body weight for 3 weeks and determined the feeding efficiency (body weight gain/food intake). The BRS-3-deficient mice at 9–10 months gained 0.028 ± 0.005 g, roughly about twice as much as the wild-type mice, which gained 0.012 ± 0.005 g ($P < 0.05$). We then analysed the rate of weight loss when food was restricted. Mice at 11–15 weeks of age were fed with 5% body weight of a standard laboratory diet. Initially, the body weight of BRS-3-deficient mice (29.7 ± 1.3 g) was similar to that of control mice (27.1 ± 0.5 g) ($P = 0.10$), but after 5 days the weight reduction in mutant mice was lower than in the controls (Fig. 2c). These results suggest that the BRS-3-deficient mice were not only hyperphagic, but also had increased feed efficiency, as a result of which they exhibited a slower rate of weight loss when food was restricted.

We then studied the energy expenditure of these mice, and began by measuring locomotor activity. When put in a cage, BRS-3-deficient mice did not show reduced activity during a test period lasting 1 h (Fig. 5a). We monitored their activity continuously over 3 weeks, and found that they had activity levels similar to those of the control mice both during light and dark periods (data not shown). To assess whether thermogenesis was reduced in mutant mice, we measured their rectal temperature (Fig. 5b). We found no defective regulation of body temperature, either at 24°C or after exposure to 4°C for 2 h, indicating that non-shivering thermogenesis (NST) was not affected in the BRS-3-deficient mice. Because NST is known to be induced by mitochondrial uncoupling protein (UCP) in brown adipose tissue, we determined the level of UCP1 mRNA in brown adipose tissue. After exposure to cold, expression of UCP1 mRNA (UCP1 mRNA density unit per μg total RNA) increased 1.6-fold compared with the level at 24°C in both control and BRS-3-deficient mice (Fig. 5c). These results demonstrate that non-

shivering brown-fat thermogenesis was not defective in the mutant mice. Because neither physical activity nor NST was affected in these animals, it remained possible that they have a lower energy expenditure resulting from a lower basal (fasting) metabolic rate and/or reduced diet-induced thermogenesis. We therefore measured oxygen consumption in these mice. For this purpose, we selected the BRS-3-deficient and control mice at 6–7 weeks of age, when body weight was not significantly different (23.2 ± 0.8 versus 22.2 ± 0.2 g, respectively; $P = 0.20$). Indeed, BRS-3-deficient mice consumed 11% less oxygen than control mice in the resting state ($P < 0.05$; Fig. 5d). This result shows that obesity in the mutant mice was preceded by a reduced metabolic rate, in accordance with their slower weight loss.

In BRS-3-deficient mice, the level of circulating leptin was elevated by 38% at 10 weeks, and at 21–22 weeks this reached a concentration five times higher than that measured in control mice. It has been shown that the leptin level reflects the mass of adipose tissue and reduction of leptin sensitivity may lead to obesity^{10–12}. Consistent with this, BRS-3-deficient mice became obese despite the elevated level of plasma leptin, suggesting that leptin sensitivity was reduced.

Our data demonstrate that one of the subtypes of the bombesin-like peptide receptors in mammals helps to regulate energy balance. BRS-3 is expressed in brain and spinal cord, but is not highly expressed in other tissues examined so far, including adipose tissue, liver, pancreas and skeletal muscle⁸ (data not shown). In brain, the expression of BRS-3 is prominent in the hypothalamic nuclei including the paraventricular, arcuate and dorsomedial nuclei, which are involved in regulation of both endocrine processes, and the autonomic nervous system. Because the gross anatomy of the hypothalamic region was not changed in the BRS-3-deficient mice, their obesity may not involve a developmental abnormality of the hypothalamus.

Our analysis of BRS-3-deficient mice demonstrates the physiological role of this orphan receptor. It will be interesting to see whether a synthetic peptide that has high affinity for both BRS-3

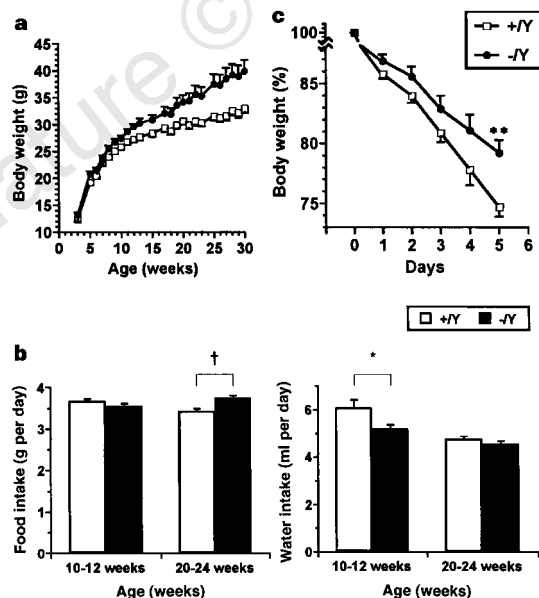


Figure 2 Intake of food and water and change in body weight in wild-type and BRS-3-deficient mice. **a**, Growth curves of wild-type (+Y) ($n = 10$) and BRS-3-deficient mice (-Y) ($n = 8$). From 8 to 18 weeks, $P < 0.05$; from 19 weeks onwards, $P < 0.02$. **b**, Food (left) and water (right) consumption in wild-type (+Y) and BRS-3-deficient mice (-Y). *, $P < 0.05$; †, $P < 0.002$. **c**, Body-weight change when food was restricted. **, $P < 0.01$. All values are means $\pm$ s.e.m. Statistics were performed with a two-tailed Student's *t*-test for **a** and **c**, and with analysis of variance (ANOVA, two-way layout with repetition) for **b**.

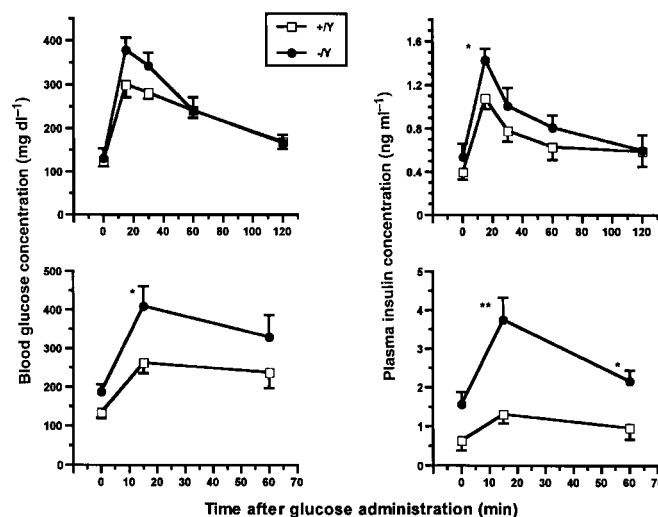


Figure 3 Oral glucose tolerance tests for wild-type and mutant mice. Wild-type (+Y) and mutant (-Y) mice are shown at 9–10 weeks (top) and 26 weeks of age (bottom). Changes in blood glucose level (left) and insulin level (right) after glucose administration are shown. Values are means $\pm$ s.e.m. Statistics were performed with a two-tailed unpaired Student's *t*-test. All *P*-values are from comparisons between BRS-3-deficient and control mice at the time indicated. *, $P < 0.05$; **, $P < 0.02$.

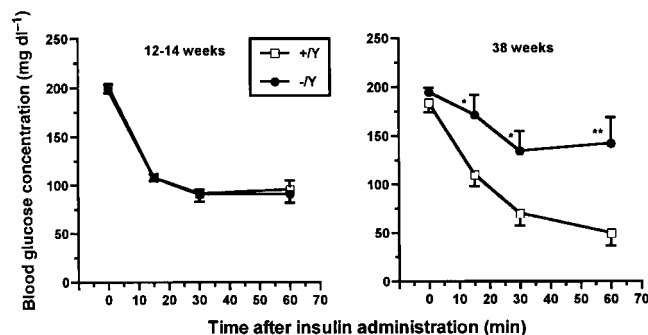


Figure 4 Insulin tolerance tests for wild-type and mutant mice. Wild type (+/Y) and mutant (-/Y) mice are shown at 12–14 weeks (left) and 38 weeks of age (right). Values are means $\pm$ s.e.m. Statistical analysis was as described in Fig. 3. *, $P < 0.05$; **, $P < 0.02$.

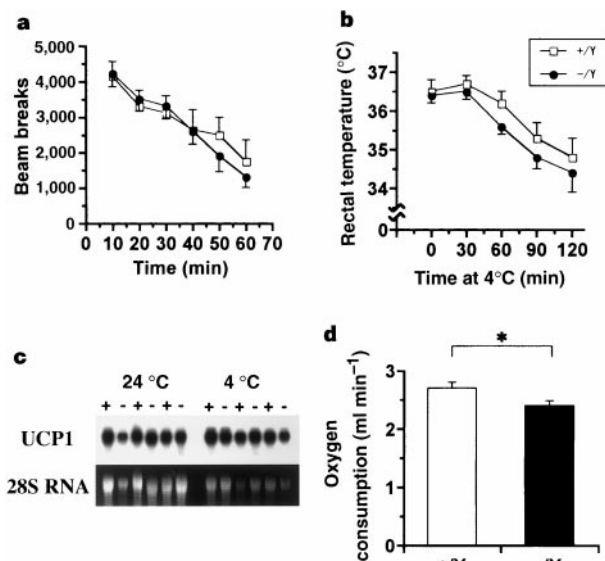


Figure 5 Measures of metabolic activity in wild-type and BRS-3-deficient mice. **a**, Locomotor activity measured by number of breaks of a beam. **b**, Thermoregulation during cold exposure. **c**, UCP1 expression in brown adipose tissue. Mice were kept either at 24°C all the time or at 4°C for 2 h, and total RNA from brown adipose tissue was isolated and analysed for UCP1 expression. **d**, Metabolic rate in mutant and wild-type mice. Oxygen consumption was measured in the resting state. Values are means $\pm$ s.e.m. Statistical analysis was as described in Fig. 3. *, $P < 0.05$.

and GRP-R¹³ has any effect on metabolism when injected into animals. Identification of the natural ligand for this receptor will be an important step that could lead to the development of therapies for obesity and associated dysfunctions.

Methods

Targeted disruption of the BRS-3 gene and generation of BRS-3-deficient mice. For constructing the targeting vector, a 9.5-kilobase *SpeI*–*SalI* fragment containing all three exons of the BRS-3 gene was cloned from a 129SV mouse genomic library, and a 1.5-kb *EcoRI*–*EcoRV* fragment containing exon 2 was inactivated by substitution with a Neo-cassette from pKJ2 (kindly provided by T. Yagi). This plasmid contains an open reading frame for a mutation-corrected *neo* gene, which is inserted in the mouse *pgk-1* gene^{14,15}. The thymidine kinase gene for negative selection (a gift from Y. Nabeshima) was ligated to the 5' position of the targeting vector¹⁶ (Fig. 1a). The E14 embryonic stem (ES) cell line¹⁷ (a gift from M. Hooper) was transfected with linearized vector by electroporation, and we obtained one homologous recombinant for 347 G418 and gancyclovir double-resistant clones screened (Fig. 1b). DNA was

digested with *EcoRI* and *Bam*HI endonuclease and restriction fragments were detected by Southern analysis using the indicated 5' probe (a *SalI*–*SpeI* fragment). To confirm the correct homologous recombination event, *SpeI*- and *EcoRV*-digested fragments were detected with 3' probe (a *SmaI*–*SmaI* fragment) and Neo probe, respectively. ES cells from this clone was injected into the blastocysts of B6, and chimaeric embryos were implanted into the uterus of pseudopregnant Jcl:ICR (Institute of Cancer Research) mice. Chimaeric male mice were then mated to female B6 mice, and germline transmission of the ES cells was confirmed. Heterozygous female mice (F₁) were then mated to B6 mice to produce F₂ male wild-type (+/Y) and hemizygous mutant mice (-/Y). We backcrossed heterozygous F₂ female mice to B6 to produce F₃ male mice, and backcrossed heterozygous F₃ female mice to B6 for F₄ production. For each analysis we used BRS-3-deficient mice and control wild-type mice of the same generation. F₂, F₃ and F₄ generations were used throughout the experiment, except for analysis of body weight, in which the (F₂ $\times$ F₂) generation also was used to confirm the obese phenotype. Genotyping of mice was also performed by PCR with the primers indicated (1, 2 and 3 in Fig. 1a). Primer sequences: wild-type primer 1, ATG CAA CCC ACT ACC TGG CAG AA; wild-type primer 2, ACA GGT CTT CAG AAT GGC ATT GG; mutant primer 3, TAC GGT ATC GCC GCT CCC GAT T.

Blood analysis. For oral glucose tolerance tests, male mice were deprived of food for 16 h and given 0.5 mg glucose per g body weight orally; 6 wild-type and 6 mutant mice at 9–10 weeks and 4 wild-type and 4 mutant mice at 26 weeks were used. For insulin tolerance tests, after a 6-h fast, male mice were given 0.75 U human insulin per kg body weight by subcutaneous injection; 5 wild-type and 6 mutant mice at 12–14 weeks, as well as 6 wild-type and 6 mutant mice at 38 weeks, were used. Blood was withdrawn from the tail at indicated times. Blood glucose levels and plasma insulin levels were measured by the glucose oxidase method (Boehringer) and by ELISA (Morinaga), respectively. Plasma levels of cholesterol and triacyl glycerol were measured with colorimetric kits (Boehringer). Commercial radioimmunoassay kits were used to measure serum glucagon (Dai-ichi RI), free T₃ and free T₄ (Kodak Clinical Diagnostics) levels. Serum growth hormone levels were determined by radioimmunoassay using purified mouse growth hormone and polyclonal monkey anti-rat growth-hormone antibody. All procedures for the assay were performed as described¹⁸.

Locomotor activity. Locomotor activity was measured using Scanet (Model SV-10, Toyo Sangyo, Toyama). Mice (12 of each genotype at 10 weeks of age) were placed individually in a plastic cage and the number of infrared beam breaks was monitored every 10 min for 60 min. Long-term measurement was made by using Animex Auto (Model MK-110, Muromachikikai, Tokyo). In this case, mice were kept in their home cage with food and water.

Physiological measurements. For measurement of food and water consumption, 6 wild-type and 6 mutant mice were housed individually. Consumption of standard mouse chow and water, as well as body weight, were measured for 4 consecutive days at indicated ages. Feeding efficiency was determined in 6 mutant and 6 control mice at 38–40 weeks. When food was restricted, 7 mutants and 6 control mice at 11–15 weeks of age were fed with 5% body weight of a standard laboratory diet. Heart rate and blood pressure were measured in conscious mice using a blood-pressure monitor (Model MK-1030, Muromachikikai, Tokyo). Rectal temperature was monitored (12 mice of each genotype at 14 weeks of age) using an electronic thermister (Model BAT-12 Physitemp, NJ) equipped with a rectal probe (RET-3 Physitemp, NJ). Oxygen consumption was determined in 6 mutant and 6 control mice of 6–7 weeks of age, with an O₂/CO₂ metabolism measuring system (Model MK-5000, Muromachikikai, Tokyo) at 24°C. The chamber volume was 560 ml, airflow to the chamber was 300 ml min⁻¹, samples were taken every 3 min and a standard gas reference was taken every 30 min. Mice were kept unrestrained in the chamber for about 3 h without food or water during the light cycle, and the oxygen consumption was determined when the minimum plateau shape was obtained, which corresponded to the period of sleep or inactivity of the mice (usually 1.5–2.5 h after they had been placed in the chamber).

Analysis of UCP1 expression. Total RNA was isolated from the brown adipose tissue of mice at 26 weeks of age. Northern blotting was performed by resolving 8 μ g of total RNA from mice kept at 24°C and 5 μ g of total RNA from mice exposed to 4°C for 2 h. The hybridization probe was rat UCP1 cDNA (position 743–1103, a gift from H. Ohno)¹⁹.

Received 5 June; accepted 18 August 1997.

1. Lebacqz-Verheyden, A.-M., Trepel, J., Sausville, E. A. & Battey, J. in *Handbook of Experimental Pharmacology*, Vol. 95/II. *Peptide Growth Factors and their Receptors II* (eds Sporn, M. B. & Roberts, A. B.) 71–124 (Springer, Berlin, Heidelberg, 1990).
2. Spindel, E. R., Giladi, E., Brehm, P., Goodman, R. H. & Segerson, T. P. Cloning and functional characterization of a complementary DNA encoding the murine fibroblast bombesin/gastrin-releasing peptide receptor. *Mol. Endocrinol.* **4**, 1956–1963 (1990).
3. Battey, J. F. *et al.* Molecular cloning of the bombesin/gastrin-releasing peptide receptor from Swiss 3T3 cells. *Proc. Natl Acad. Sci. USA* **88**, 395–399 (1991).
4. Corjay, M. H. *et al.* Two distinct bombesin receptor subtypes are expressed and functional in human lung carcinoma cells. *J. Biol. Chem.* **266**, 18771–18779 (1991).
5. Wada, E. *et al.* cDNA cloning, characterization, and brain region-specific expression of a neuromedin-B preferring bombesin receptor. *Neuron* **6**, 421–430 (1991).
6. Fathi, Z. *et al.* A novel bombesin receptor subtype selectively expressed in testis and lung carcinoma cells. *J. Biol. Chem.* **268**, 5979–5984 (1993).
7. Gorbulev, V., Akhundova, A., Buchner, H. & Fahrenholz, F. Molecular cloning of a new bombesin receptor subtype expressed in uterus during pregnancy. *Eur. J. Biochem.* **208**, 405–410 (1992).
8. Ohki-Hamazaki, H., Wada, E., Matsui, K. & Wada, K. Cloning and expression of the neuromedin B receptor and the third subtype of bombesin receptor genes in the mouse. *Brain Res.* **762**, 165–172 (1997).
9. Bray, G. A. & York, D. A. Hypothalamic and genetic obesity in experimental animals: An autonomic and endocrine hypothesis. *Physiol. Rev.* **59**, 719–809 (1979).
10. Frederich, R. C. *et al.* Leptin levels reflect body lipid content in mice: Evidence for diet-induced resistance to leptin action. *Nature Med.* **1**, 1311–1314 (1995).
11. Maffei, M. *et al.* Leptin levels in human and rodent: Measurement of plasma leptin and *ob* RNA in obese and weight-reduced subjects. *Nature Med.* **1**, 1155–1161 (1995).
12. Schwartz, M. W., Peskind, E., Raskind, M., Boyko, E. J. & Porte, D. Jr Cerebrospinal fluid leptin levels: Relationship to plasma levels and to adiposity in humans. *Nature Med.* **2**, 589–593 (1996).
13. Wu, J. M., Nitecki, D. E., Biancalana, S. & Feldman, R. I. Discovery of high affinity bombesin receptor subtype 3 agonists. *Mol. Pharmacol.* **50**, 1355–1363 (1996).
14. Boer, P. H. *et al.* Polymorphisms in the coding and noncoding region of murine *Pgk-1* alleles. *Biochem. Gen.* **28**, 299–308 (1990).
15. Yanofsky, R. L., Fine, M. & Pellow, J. W. A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure. *Proc. Natl Acad. Sci. USA* **87**, 3435–3439 (1990).
16. Nabeshima, Y. *et al.* Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364**, 532–535 (1993).
17. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292–295 (1987).
18. Yamamoto, K. & Kikuyama, S. Radioimmunoassay of prolactin in plasma of bullfrog tadpoles. *Endocrinol. Jpn* **29**, 159–167 (1982).
19. Bouillaud, F., Weissenbach, J. & Ricquier, D. Complete cDNA-derived amino acid sequence of rat brown fat uncoupling protein. *J. Biol. Chem.* **261**, 1487–1490 (1986).

Acknowledgements. We thank J.-I. Miyazaki for ES cells and training regarding ES cells; A. F. Parlow for mouse growth hormone and polyclonal monkey anti-rat GH serum; K. Wakabayashi for goat anti-mouse IgG serum; H. Ohno for rat UCP1 probe; T. Nishikawa for providing Animex Auto; N. M. Le Douarin, K. Mikoshiba, R. S. Petralia and J. Smith for critical reading of the manuscript. This work was supported in part by research grants from the Ministry of Education, Science, Sports and Culture, the Ministry of Health and Welfare, the Science and Technology Agency of Japan, the Japan Health Science Foundation.

Correspondence and requests for materials should be addressed to H.O.-H. (e-mail: hamazaki@prit.go.jp) or K.W. (e-mail: wada@ncnaxp.ncnp.go.jp).

Math1 is essential for genesis of cerebellar granule neurons

Nissim Ben-Arie*†, Hugo J. Bellen*‡, Dawna L. Armstrong§, Alanna E. McCall†, Polina R. Gordadze†, Qiuxia Guo§, Martin M. Matzuk*§ & Huda Y. Zoghbi*†||

Departments of *Molecular and Human Genetics, †Pediatrics, ‡Cell Biology, §Pathology, and ||Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

The cerebellum is essential for fine motor control of movement and posture, and its dysfunction disrupts balance and impairs control of speech, limb and eye movements. The developing cerebellum consists mainly of three types of neuronal cells: granule cells in the external germinal layer, Purkinje cells, and neurons of the deep nuclei¹. The molecular mechanisms that underlie the specific determination and the differentiation of each of these neuronal subtypes are unknown. *Math1* (refs 2, 3), the mouse homologue of the *Drosophila* gene *atonal*⁴, encodes a basic helix–loop–helix transcription factor that is specifically expressed in the precursors of the external germinal layer and their derivatives. Here we report that mice lacking *Math1* fail to form granule cells and are born with a cerebellum that is devoid of an external germinal layer. To our knowledge, *Math1* is the first

gene to be shown to be required *in vivo* for the genesis of granule cells, and hence the predominant neuronal population in the cerebellum.

The mammalian cerebellum consists of deep centrally located neurons, referred to as the deep nuclei, and a peripheral cortex. The cortex contains two principal neuronal subtypes, the Purkinje cells and the cells of the internal granular layer (IGL). The IGL neurons, which eventually constitute most of the neurons in the cerebellum, are derived from peripherally located cells in the external germinal layer (EGL) which migrate postnatally along radial glia^{1,5}. The mouse cerebellar anlage is specified at embryonic-day 9 (E9) just anterior to the area where closure of the neural tube is incomplete. The neuroepithelium of the ventricular zone in the metencephalon gives rise to the precursors of the neurons of the deep nuclei and Purkinje cells⁶. The precursors of the EGL are derived at E13–E15 from a structure named the rhombic lip (also known as the germinal trigone), which is located at the posterior edge of the cerebellar anlage^{1,5,7} (Fig. 1). The proliferating cells of the rhombic lip, which are committed to become EGL neurons, disperse rostrally over the surface of the cerebellar anlage, where they establish the EGL. Proliferation of these granule cell precursors continues in the EGL until postnatal day 15 (P15). The EGL cells migrate inwardly to give rise to the IGL, starting at birth until day P20, when maturation of the cerebellum is complete^{1,5,7}.

The *Drosophila atonal* gene is essential for the generation of the chordotonal organs, the proprioceptors, and the R8 photoreceptor precursors in the eye^{4,8}. Adult flies that lack *atonal* function do not have photoreceptors, are uncoordinated and tend not to fly⁸. Expression of *Math1*, the mouse *atonal* homologue, is highly restricted to a few neurons in the hindbrain and dorsal neural tube^{2,3}. *Math1* is expressed in the precursors of the EGL in the rhombic lip starting at E13, and later in the EGL of the developing cerebellum (Fig. 1). *Math1* expression ceases during the inward migration phase of the granule cells to the IGL. To investigate the role of *Math1* in cerebellar development, a targeted deletion of the *Math1* gene (*Math1*^{ml}) was generated by using embryonic stem (ES) cell technology (Fig. 2a). Heterozygous mice were intercrossed to obtain mice homozygous for the *Math1* deletion (Fig. 2b). Absence of a hybridization signal with DNA from homozygous mutant mice, when probed with an internal probe spanning the basic-helix–

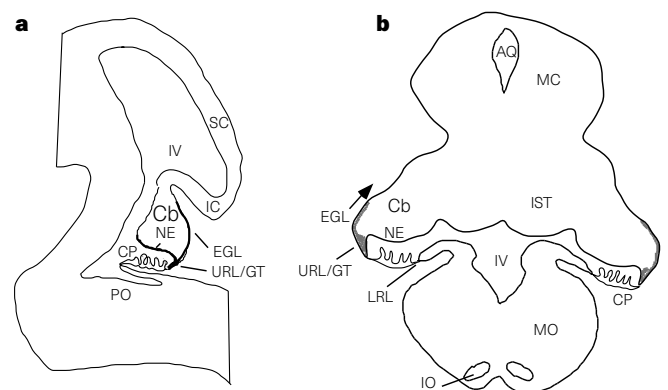


Figure 1 The developing cerebellum at E16.5. The sagittal (a) and coronal (b) views show the localization of the cerebellum (Cb) between the inferior colliculus (IC), pons (PO), mesencephalon (MC) and medulla oblongata (MO). The germinal matrix, which gives rise to EGL precursors, is the rhombic lip, with an upper and a lower component (URL, LRL, respectively). The proliferative zone at the URL, the germinal trigone (GT), has three prongs: EGL, neuroepithelium (NE), and choroid plexus (CP). Granule cell precursors originate at the URL/GT and migrate (arrow) to form the EGL. The LRL will contribute to the medullary precerebellar nuclei, including the inferior olive (IO). *Math1* is expressed in granule precursors in GT and in emerging EGL (shaded area). AQ, aqueduct of Sylvius; IST, isthmus; IV, fourth ventricle; SC, superior colliculus. Adapted from ref. 1.

loop-helix (bHLH) domain, confirmed that the mutation is a null allele (data not shown).

Mice heterozygous for the *Math1* deletion are viable, fertile and appear normal. Genotyping of newborn mice from heterozygous intercrosses demonstrated that the targeted allele did not result in embryonic lethality. The observed ratio of genotypes fitted that expected for a fully penetrant, mendelian recessive inheritance pattern, with 28 (25.2%) wild types, 55 heterozygote (49.5%) and 28 (25.2%) homozygote (null) mice. *Math1*-null mice are born alive, but fail to breathe, become cyanotic, and die a few minutes after birth. We observed no morphological abnormalities in the lung (data not shown). Furthermore, the lack of expression of *Math1* in the lungs, combined with the expression pattern in the brainstem², suggest a central mechanism for this respiratory failure.

The structure and histology of the cerebellum of *Math1*-null mice is abnormal compared to their littermates. The most peripheral cell layer in the cerebellum, the EGL, is normally about eight cells thick at E18.5. These rapidly proliferating cells are the granule cell precursors and express *Math1* abundantly². *Math1*-null mice completely lack the EGL, as shown by haematoxylin-and-eosin staining (Fig. 3a, b), and cresyl violet staining (data not shown), and have a cerebellum that is reduced in size. To analyse the changes in the cytoarchitecture of the cerebellum, we stained the two principal neuronal populations in the cerebellum, the Purkinje cells (Fig. 3c, d) and the granule cells (Fig. 3e, f). RNA *in situ* hybridization using RU49, a zinc-finger transcription factor that labels the EGL⁹, demonstrates the absence of the EGL and migrating granule cells in the cerebella of *Math1*-null mice (Fig. 3e, f). In contrast to the lack of granule neurons, the Purkinje cells and the neurons of the deep nuclei are still present in mutant mice, as demonstrated by staining with anti-calbindin antibody (Fig. 3c, d) and Nissl staining (data not shown). However, because of the absence of the EGL, the Purkinje cells, which are normally located beneath the EGL, are now localized at the periphery of the cerebellum and do not form a distinct and organized layer. In addition, a significant subpopulation of Purkinje cells fails to migrate from the central area of the cerebellum into the periphery (Fig. 3c, d). Staining with glial fibrillary acidic protein¹⁰ (GFAP) and nestin¹¹, which both label glial cells, demonstrates that there is not excessive gliosis, and that the radial glia are localized properly (data not shown). These results show that *Math1* is essential for the formation of the EGL but not the Purkinje cells, the neurons of the deep nuclei, or the glial cells. In *Math1*-null mice, the absence of the EGL also leads to the lack of foliation of the cerebellum typically observed in normal embryos starting from E18 (Fig. 3a, b).

Possible causes for the missing EGL in mutant mice include failure in fate determination of granule precursors at the rhombic lip, aberrant migration of precursors to form the EGL, or degeneration of the neurons already localized at the EGL. The granule neuron precursors originate at E13–E15 in the proliferation zone of the rhombic lip^{1–7}. At E14, the EGL precursors initiate migration from the rhombic lip over the surface of the cerebellar primordia. At this stage, *Math1* is expressed in the granule precursors localized at the rhombic lip and in the emerging EGL². As shown in Fig. 4a–d, *Math1*-deficient mice contain fewer cells in the rhombic lip than do control embryos. In addition, there is no migration of cells out of the rhombic lip to form the EGL. Bromodeoxyuridine (BrdU) labelling demonstrates the absence of proliferating cells in both sites that normally expand the pool of granule cells: the rhombic lip and the cells that have already migrated to the EGL (Fig. 4e, f). Hence, the lack of the cerebellar EGL in *Math1*-null mice is evident early on and is caused either by the absence of the granule cell precursors or by the inability of these cells to proliferate at the rhombic lip. Thus, the agenesis of granule cell precursors in *Math1*-null mice is a much earlier event than the postnatal apoptotic degeneration of the EGL seen in *weaver* mice^{12,13}.

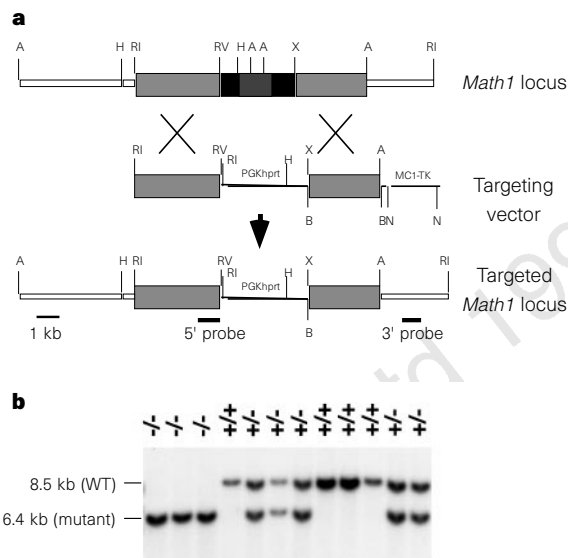


Figure 2 Targeted disruption of the *Math1* locus and generation of *Math1*-null mice. **a**, Homologous recombination between the 3.7-kb and 3.1-kb fragments in the targeting vector and the genomic locus, resulted in the replacement of 3.2 kb (black box) which include the entire coding region (hatched box). **b**, Tail DNA from newborn littermates obtained by heterozygous intercrosses was digested with *EcoRI* and *HindIII* and hybridized with the 3' diagnostic probe. The three pups homozygous for the deletion died a few minutes after delivery, whereas the four wild-type (WT) and five mice heterozygous for the deletion were viable.

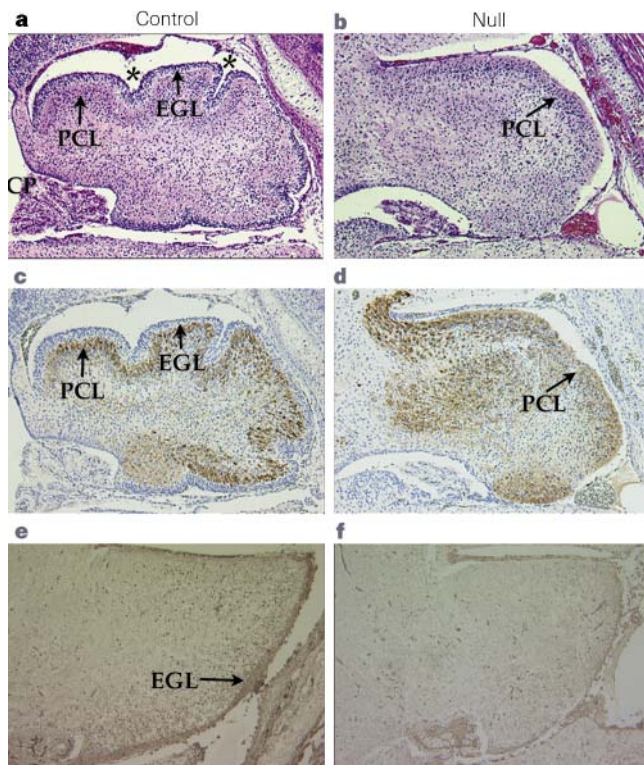


Figure 3 Cerebellar abnormalities in E18.5 *Math1*-null embryos. **a, b**, Mid-sagittal sections through the cerebellum stained with haematoxylin and eosin. The cerebellum in wild-type (**a**) mice is well developed and foliated (asterisks), but is smooth in *Math1* null mice (**b**). The external germinal layer (EGL) is absent and the Purkinje cell layer (PCL) less organized, with some PC located deep in the cerebellum, as revealed by anti-calbindin staining (**c, d**). Parasagittal sections of the cerebellum of wild-type (**e**) and mutant (**f**) embryos were hybridized with the granule-cell marker RU49. Wild-type EGL is strongly positive for RU49, whereas the cerebellum of *Math1*-null mice is negative. Original magnifications: **a–d**, $\times 100$; **e, f**, $\times 40$.

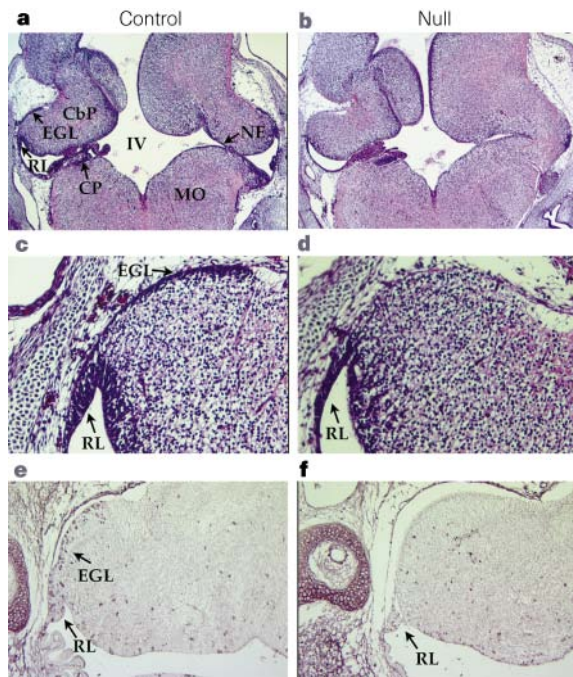


Figure 4 Absence of the EGL in E14.5 *Math1* null mice due to lack of proliferation of granule precursors. **a–d**, Generation of the EGL by cells migrating from the rhombic lip (haematoxylin and eosin stain). In coronal sections from wild-type brain (**a, c**), the EGL is identified as cells migrating out from the rhombic lip (RL) over the cerebellar primordium (CbP). In *Math1* null mice (**b, d**), the rhombic lip is thinner and lacks EGL precursors migrating rostrally. **e, f**, The absence of the EGL precursors in *Math1* null mice is supported by the lack of proliferating cells in the rhombic lip of these mice, as demonstrated by BrdU staining. Proliferating cells are seen at the rhombic lip and EGL of wild-type mice (**e**), but not in *Math1* null mice (**f**). Original magnifications: **a, b**, $\times 40$; **c, d**, $\times 200$; **e, f**, $\times 100$.

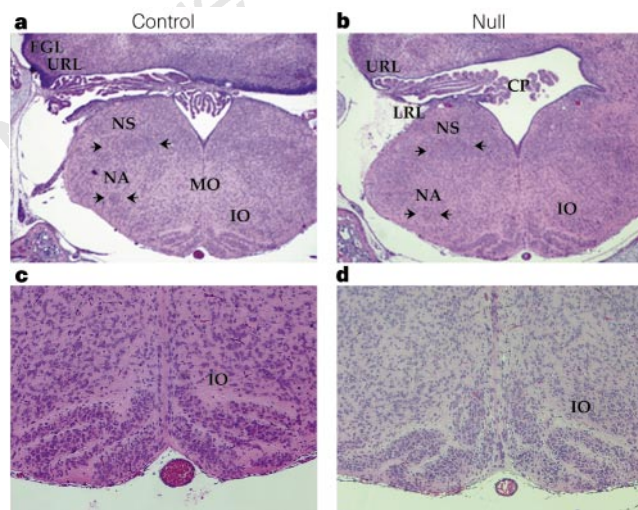


Figure 5 Inferior olive nuclei are normal in *Math1* null mice. Coronal sections from brains of *Math1* heterozygous (**a, c**) and null mice (**b, d**) at E18.5 were stained with haematoxylin and eosin. The inferior olive (IO) nuclei in the medulla oblongata (MO), which arise from the lower rhombic lip (LRL) appear normal in both animals (**a–d**). Arrows indicate the nucleus tractus solitarius (NS) and nucleus ambiguus (NA), which appear normal in both animals (**a, b**). This is in contrast to the lack of the external germinal layer (EGL), and the smaller upper rhombic lip (URL) seen in *Math1* null mice (**b**). Original magnifications: **a, b**, $\times 40$; **c, d**, $\times 100$.

It has been suggested that inferior olive neurons are derived from the rhombic lip¹⁴, which consists of an upper and a lower component (Fig. 1) and that it is the lower lip that gives rise to the neurons of the inferior olive¹. *Math1* is only expressed in the germinal trigone that is part of the upper lip and not the lower lip² (Fig. 1). The inferior olive neurons are present in *Math1*-null mice and appear normal (Fig. 5), consistent with *Math1* expression in the upper rhombic lip. This finding supports the histological and functional division of the rhombic lip into an upper and a lower lip, each contributing to the generation of different types of neurons that migrate to separate destinations.

As *Math1*-null mice die shortly after birth from respiratory failure and *Math1* is expressed in the hindbrain, we wanted to establish whether there is selective loss of neurons implicated in the control of respiration, including those of the nucleus tractus solitarius, nucleus ambiguus, dorsal vagus nucleus and trigeminal ganglion. These brainstem nuclei were identified and analyzed for neuronal density and morphology. No differences between *Math1*-null mice and littermate controls were noted (Fig. 5). The lack of morphological defects suggests that *Math1* has different roles in the brainstem and cerebellum, or that *Math1* is required in other unidentified neuronal cells involved in respiratory control.

We have shown that the *Math1* gene is required for generation of the EGL cells and their derivatives. Although ~ 30 mutations in mice cause cerebellar defects and ataxia^{15–17}, none of these mutations causes a phenotype like that of *Math1*-null mice. For example, the developmental mutant *reeler* causes a defect in cell migration in the cerebellum, cerebral cortex and hippocampus, rather than a defect in cell specification or proliferation¹⁸. In the cerebellum of *weaver* mutant mice, proliferation of EGL precursor cells is normal, but failure of granule cell migration from the EGL to the IGL causes degeneration of this neuronal population¹⁹.

The cerebellum is derived from the posterior end of the mesencephalon and from the metencephalon. The expression of *wnt-1* at the mesencephalon–metencephalon junction stabilizes the expression of the *engrailed* homologues *En-1* and *En-2*, thereby specifying the cerebellar anlage²⁰. Absence of *Wnt-1* or of *En-1* and *En-2* causes a loss of cerebellar components, including the deep nuclei, Purkinje cells and IGL^{21–26}. In the absence of *Math1*, the cerebellum develops but there is selective loss of only the EGL precursors and granule neurons. Other defects, like the ectopic localization of the Purkinje cells and the lack of foliation are most likely to be secondary to the absence of EGL. Mice lacking *Mash1*, the murine homologue of the *Drosophila* proneural gene complex *achaete-scute*, suffer from loss of olfactory neuronal progenitors and are arrested in the differentiation of sympathetic neuronal precursors²⁷. Therefore, *Mash1* and *Math1* are both essential for the generation of neurons through progression of a differentiation programme, possibly in selected progenitors. On the basis of our results, we propose that *Math1* plays a role in cerebellar granule fate specification or proliferation. This role is functionally similar, although not identical, to the role of its homologue, the *Drosophila atonal* gene^{4,8}. In both mice and fruitflies, loss of function leads to the absence of specific neuronal cell types early in neurogenesis. However, *Math1* is expressed in cells that are already committed to be neural tissue, whereas *atonal* plays a role in the decision of epidermal rather than neuronal cell fate. □

Methods

Targeted deletion of *Math1*. The genomic locus was targeted by homologous recombination using two isogenic flanking fragments of 3.7 kb (*EcoRI*–*EcoRV*) and 3.1 kb (*XbaI*–*ApaI*), which were isolated from a mouse 129/SvEv genomic DNA library (Stratagene). A human hypoxanthine–guanine phosphoribosyl-transferase (*hprt*) and the herpes simplex virus thymidine kinase gene were used as positive and negative selection markers, respectively. The linearized targeting construct was electroporated into *hprt*-negative AB2.1 embryonic stem (ES) cells and selection was achieved using HAT (hypoxanthine,

aminopterin and thymidine) and FIAU (1-(2'-deoxy-2'-fluoro- β -D-arabino-furanosyl)-5-iodouracil). The targeted alleles can be identified by the presence of a 14.6-kb fragment instead of 10 kb, when DNA is digested with *Apal* and probed with a 5' diagnostic probe, and by a fragment of 6.4 kb instead of 8.5 kb, when digested with *EcoRI* and *HindIII* and probed with an external 3' probe. Homologous recombination was detected in 74% (130/175) of the ES clones screened. Two independent ES cell lines heterozygous for the targeted *Math1* allele (AT37-G6 and AT37-G12) were injected into blastocysts and gave rise to chimaeric mice, which transmitted the deletion through the germline.

Histology, immunohistochemistry and *in situ* hybridization. Pregnant females were killed and embryos removed quickly into cold PBS. Extra-embryonic membranes were discarded, and the yolk sac was collected for DNA extraction. Embryos or newborn mice were fixed overnight in either Bouin's fixative or 4% paraformaldehyde in PBS at 4 °C, dehydrated and stored at -20 °C. Embryos were embedded in paraffin wax, and sections of 7 μ m were cut. Sections were stained in Mayer's haematoxylin and eosin or cresyl violet. Isotopic *in situ* hybridization with RNA probes labelled with [³⁵S]UTP has been described²⁸.

Non-radioactive *in situ* hybridization using the digoxigenin system (Boehringer Mannheim) was done according to ref. 29, but with paraffin embedding. Sections were dewaxed (HistoClear, National Diagnostics), rehydrated through a graded series of ethanol concentrations, and processed according to the protocol. Templates for cRNA synthesis were as follows: *Math1* (PCR-amplified coding region) and RU49 (0.847-kb *EcoRI/PstI* fragment, a gift from N. Heintz). For immunohistochemical staining, paraffin blocks were cut at 5 μ m, dewaxed in xylene, antigen-retrieved by microwaving in citrate buffer, blocked and incubated with the appropriate antibodies in PBS containing 10% normal goat serum for several days at 4 °C. Primary antibodies were detected using the avidin-biotin complex system (ABC, Vector Lab). Antibodies used were as follows: anti-calbindin-D (1:1,000; Sigma), anti-GFAP (1:1,000; Dako) anti-*nestin* (1:150, a gift from R. McKay).

For evaluation of brainstem cranial nerve nuclei, coronal sections (6 μ m) were cut from E18.5 fetuses and alternative sections were stained with haematoxylin and eosin, cresyl violet (Nissl staining) and one-step trichrome (Gomori's staining). Cranial nerve nuclei were identified and scored independently by two individuals. Criteria for the examination were: localization of each cranial nerve nucleus, symmetry, shape and size, and neuronal density.

BrdU staining of proliferating cells. Pregnant females were injected with BrdU (5-bromodeoxyuridine) (50 mg kg⁻¹ body weight from a stock solution of 10 mg ml⁻¹ in 0.1M Tris-HCl, pH 7.5) and killed two hours later. Embryos were dissected in cold PBS, fixed in 70% ethanol, dehydrated, embedded in paraffin wax and cut at 7- μ m sections. These were treated with 2M HCl for 30 min at 37 °C to denature the DNA partially, then neutralized in 0.1M sodium borate, pH 8.5, for 10 min, as recommended by the supplier of the anti-BrdU antibodies (Novocastra Lab). BrdU-labelled cells were detected using the ABC system.

Received 19 May; accepted 7 August 1997.

- Altman, J. & Bayer, S. A. in *Development of the Cerebellar System: In Relation to its Evolution, Structure, and Functions* (CRC Press, Boca Raton, Florida, 1996).
- Akazawa, C., Ishibashi, M., Shimizu, C., Nakanishi, S. & Kageyama, R. A. Mammalian helix-loop-helix factor structurally related to the product of *Drosophila* proneural gene *atonal* is a positive transcriptional regulator expressed in the developing nervous system. *J. Biol. Chem.* **270**, 8730-8738 (1995).
- Ben-Arie, N. et al. Evolutionary conservation of sequence and expression of the bHLH protein *Atonal* suggest a conserved role in neurogenesis. *Hum. Mol. Genet.* **5**, 1207-1216 (1996).
- Jarman, A. P., Grau, Y., Jan, L. Y. & Jan, Y. N. *Atonal* is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system. *Cell* **73**, 1307-1321 (1993).
- Hatten, M. E. & Heintz, N. Mechanisms of neural patterning and specification in the developing cerebellum. *Annu. Rev. Neurosci.* **18**, 385-408 (1995).
- Hallonet, M. E. & Le Dourain, N. M. Tracing neuroepithelial cells of the mesencephalic and metencephalic alar plates during cerebellar ontogeny in quail-chick chimaeras. *Eur. J. Neurosci.* **5**, 1145-1155 (1993).
- Hatten, M. E., Alder, J., Zimmerman, K. & Heintz, N. Genes involved in cerebellar cell specification and differentiation. *Curr. Opin. Neurobiol.* **7**, 40-47 (1997).
- Jarman, A. P., Sun, Y., Jan, L. Y. & Jan, Y. N. Role of the proneural gene, *atonal*, in formation of *Drosophila* chordotonal organs and photoreceptors. *Development* **121**, 2019-2030 (1995).
- Yang, X. W., Zhong, R. & Heintz, N. Granule cell specification in the developing mouse brain as defined by expression of the zinc finger transcription factor RU49. *Development* **122**, 555-566 (1996).
- Debus, E., Weber, K. & Osborn, M. Monoclonal antibodies specific for glial fibrillary acidic (GFA) protein and for each of the neurofilament triplet polypeptides. *Differentiation* **25**, 193-203 (1983).
- Tohyama, T. et al. *Nestin* expression in embryonic human neuroepithelium and in human neuroepithelial tumor cells. *Lab. Invest.* **66**, 303-313 (1992).
- Smeyne, R. J. & Goldowitz, D. Development and death of external granular layer cells in the *weaver* mouse cerebellum: a quantitative study. *J. Neurosci.* **9**, 1608-1620 (1989).

- Wood, K. A., Dipsasquale, B. & Youle, R. J. *In situ* labeling of granule cells for apoptosis associated DNA fragmentation reveals different mechanisms of cell loss in developing cerebellum. *Neuron* **11**, 621-632 (1993).
- His, W. Die Entwicklung des menschlichen Rautenhirns vom Ende des ersten bis zum Beginn des dritten Monats. I. Verlängertes Mark. *Abhandlung der königlichen sächsischen Gesellschaft der Wissenschaften, mathematische-physikalische Klasse* **29**, 1-74 (1891).
- Heintz, N., Norman, D., Gao, W.-O. & Hatten, M. in *genome Maps and Neurological Disorders* (eds Davies, K. E. & Tilghman, S. M.) 19-45 (Cold Spring Harbor Laboratory Press, New York, 1993).
- Mouse Genome Database 3.1 (Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine (<http://www.informatics.jax.org/>), 1996).
- Frontiers in Bioscience 2.6 (<http://www.bioscience.org/knockout/knohome.html>) (1997).
- Goffinet, A. M. Events governing organization of postmigratory neurons: studies on brain development in normal and *reeler* mice. *Brain Res. Rev.* **7**, 261-296 (1984).
- Rezaei, Z. & Yoon, C. H. Abnormal rate of granule cell migration in the cerebellum of *weaver* mutant mice. *Dev. Biol.* **29**, 17-26 (1972).
- Joyner, A. L. *Engrailed*, *Wnt* and *Pax* genes regulate midbrain-hindbrain development. *Trends Genet.* **12**, 15-20 (1996).
- McMahon, A. P. & Bradley, A. The *Wnt-1* (*int-1*) proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62**, 1073-1085 (1990).
- Thomas, K. R. & Capecchi, M. R. Targeted disruption of the murine *int-1* proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. *Nature* **346**, 845-850 (1990).
- Thomas, K. R., Musci, T. S., Neumann, P. E. & Capecchi, M. R. *Swaying* is a mutant allele of the proto-oncogene *Wnt-1*. *Cell* **67**, 969-976 (1991).
- McMahon, A. P., Joyner, A. L., Bradley, A. & McMahon, J. A. The midbrain-hindbrain phenotype of *Wnt-1/Wnt-1* mice results from stepwise deletion of *engrailed*-expressing cells by 9.5 days postcoitum. *Cell* **69**, 581-595 (1992).
- Millen, K. J., Wurst, W., Herrup, K. & Joyner, A. L. Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse *Engrailed-2* mutants. *Development* **120**, 695-706 (1994).
- Wurst, W., Auerbach, A. B. & Joyner, A. L. Multiple developmental defects in *Engrailed-1* mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development* **120**, 2065-2075 (1994).
- Guillemot, F. et al. Mammalian *achaete-scute* homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell* **75**, 463-476 (1993).
- Albrecht, U., Eichele, G., Helms, J. & Lu, H. in *Visualization of Gene Expression Patterns by in situ Hybridization* (ed. Daston, G. P.) 23-48 (CRC Press, Boca Raton, Florida, 1997).
- Schaeren-Wiemers, N. & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: *in situ* hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100**, 431-440 (1993).

Acknowledgements. We thank B. A. Antalfy for technical assistance; P. R. Cox and J. Dong for their help; and R. R. Behringer, K. A. Mahon and B. Hassan for critical reading of the manuscript. H.Y.Z. is an investigator and H.J.B. is an associate investigator of the Howard Hughes Medical Institute. This work was supported by grants from the NIH/NINDS and by the Baylor Mental Retardation Research Center.

Correspondence and requests for materials should be addressed to H.Y.Z. (e-mail: hzhgbbi@bcm.tmc.edu).

Impaired mast cell-dependent natural immunity in complement C3-deficient mice

Andrey P. Prodeus*, Xiaoning Zhou*, Marcus Maurer†, Stephen J. Galli*† & Michael C. Carroll*

* Departments of Pathology, Harvard Medical School and † Beth Israel Deaconess Medical Center, Boston, Massachusetts, 02115, USA

The complement system is widely regarded as essential for normal inflammation, not least because of its ability to activate mast cells¹⁻⁵. However, recent studies have called into question the importance of complement in several examples of mast cell-dependent inflammatory responses⁶⁻⁹. To investigate the role of complement in mast cell-dependent natural immunity, we examined the responses of complement-deficient mice^{10,11} to caecal ligation and puncture¹², a model of acute septic peritonitis^{12,13} that is dependent on mast cells and tumour necrosis factor- α (TNF- α). We found that C4- or C3-deficient mice^{10,11} were much more sensitive to caecal ligation and puncture than wild-type (WT) controls (100% versus 20% in 24-h mortality, respectively). C3-deficient mice also exhibited reductions in peritoneal mast cell degranulation, production of TNF- α , neutrophil infiltration and clearance of bacteria. Treating the C3-deficient mice with purified C3 protein enhanced activation of peritoneal mast cells, TNF- α production, neutrophil recruitment, opsonophagocytosis of bacteria and resistance to caecal ligation and puncture, confirming that the defects were complement-dependent. These results provide formal evidence that complement activation is essential for

the full expression of innate immunity in this mast cell-dependent model of bacterial infection.

Caecal ligation and puncture (CLP) produced substantially ($P < 0.0001$) more mortality in C4-/- mice¹⁰ or complement C3 (C3-/-) mice¹¹ than in wild-type mice, particularly within the first 24 h following the procedure (100 versus 20% mortality, respectively) (Fig. 1). The enhanced susceptibility of the C3-/- mice to CLP was due to the absence of functional C3, as reconstitution of the C3-deficient animals with a single intraperitoneal injection of 0.5 mg purified human C3 protein (HuC3) reduced the 24-h mortality from 100 to 40% (Fig. 1). Previous studies had demonstrated that human C3 could be substituted for mouse C3 *in vitro*¹¹ and *in vivo*¹⁴ and that this dose, which results in blood levels in C3-/- mice that are roughly one-quarter of that in the blood of wild-type mice, was rapidly absorbed into the circulation following intraperitoneal (i.p.) injection (our unpublished results). Protection by C3 protein replacement was temporary, however, as by 48 h after CLP mortality in the HuC3-treated C3-/- mice had increased to 60%. This result was not unexpected because exogenous C3 is rapidly catabolized in the C3-deficient mice (our unpublished results).

Peritoneal mast cells recovered from the C3-deficient mice 1 or 3 h after CLP exhibited evidence of degranulation, but significantly less than that observed in the wild-type mice (Fig. 2). Thus, by 1 h after CLP, more than 75% of peritoneal mast cells identified in cytopins from wild-type mice were moderately or extensively degranulated compared to fewer than 50% of the mast cells in the C3-/- mice ($P < 0.001$). By contrast peritoneal mast cells in HuC3-reconstituted C3-/- mice exhibited levels of degranulation after CLP that were similar to those of mast cells in the wild-type mice (Fig. 2).

C3-/- ($n = 10$) and wild-type ($n = 9$) mice had similar numbers of total peritoneal cells (4.0 ± 0.2 versus $4.0 \pm 0.3 \times 10^6$ per mouse, $P > 0.05$) and total peritoneal mast cells (3.3 ± 0.5 versus $4.7 \pm 0.7 \times 10^4$ per mouse, $P > 0.05$) at baseline before CLP. However, the number of peritoneal mast cells that can be recovered by peritoneal lavage after CLP is less than can be obtained from the unmanipulated peritoneal cavity. Moreover, mast cell cytokine production does not always correlate with degranulation¹⁵. We therefore sought to use another, more relevant index of mast cell activation in this model. One hallmark of mast cell activation is the rapid release of TNF- α ^{16,17}, the cytokine that has been implicated as an important mediator in mast cell-dependent models of natural immunity, including CLP¹³ and other examples of bacterial infection¹⁸. When we measured the levels of TNF- α in peritoneal lavage fluids that were collected after CLP and assayed by enzyme-linked immunosorbent assay (ELISA), we found significantly ($P < 0.0001$) lower TNF- α levels at both 1 h (158 ± 24 pg ml⁻¹ ($n = 13$) versus 400 ± 45 pg ml⁻¹ ($n = 14$) and 3 h (218 ± 35 pg ml⁻¹ ($n = 13$) versus 663 ± 53 pg ml⁻¹ ($n = 14$)) in the C3-/- versus wild-type animals (Fig. 3a). Reconstitution with C3 protein restored TNF- α levels to those of wild-type mice, at both 1 h (344 ± 40 pg ml⁻¹ ($n = 5$) and 3 h (521 ± 61 pg ml⁻¹ ($n = 5$)) following CLP (Fig. 3a).

Neutrophils represent less than 1% of the total cells in the peritoneal cavity of untreated C3-/- or wild-type mice. In wild-type mice, neutrophils account for 50% of peritoneal cells at 1 h and 90% of cells at 3 h following CLP (Fig. 3b). In contrast, significantly ($P < 0.0001$) fewer neutrophils were found in the peritoneal cavity of C3-/- mice at the same time points after CLP (1% and 45%, respectively). TNF- α importantly contributes to mast cell-dependent neutrophil recruitment, as pretreatment of mice with an antibody specific to TNF- α can substantially reduce mast cell-dependent infiltration of neutrophils, both in the skin at sites of IgE and antigen challenge¹⁹ and in the peritoneal cavity in models of bacterial infection¹⁸ or immune complex deposition^{20,21}. Enhancement of neutrophil infiltration by TNF- α in these settings may reflect a number of different actions of this cytokine, such as

upregulation of E selection on venular endothelial cells, induction of neutrophil chemo-attractants such as interleukin-8 (IL-8), and the direct upregulation and activation of integrins on leukocytes¹⁹⁻²².

As assessed in cytocentrifuge preparations, the delayed and reduced infiltration of neutrophils into the peritoneal cavity of C3-/- mice after CLP was associated with a less efficient clearance of bacteria than in wild-type mice, by either 1 or 3 h (Fig. 4a-e). Because cytocentrifuge preparations of each of the CLP-treated C3-/- mice revealed reduced numbers of neutrophils and large numbers of bacteria (Fig. 4b-e), whereas

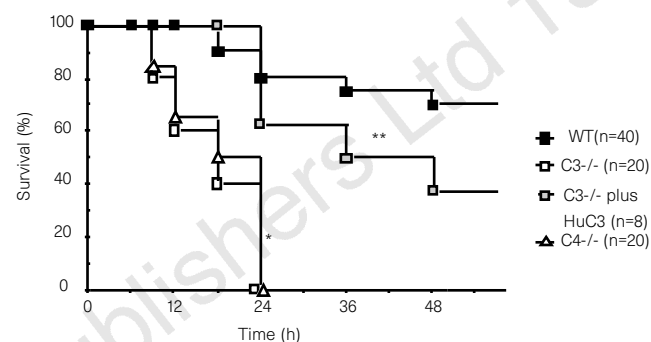


Figure 1 C3- or C4-deficient mice are highly sensitive to peritoneal bacterial infection. Acute septic peritonitis was induced in wild-type (WT) and C3- or C4-deficient mice by performing caecal ligation and puncture. Reconstitution of C3-deficient mice with 0.5 mg human C3 (administered intraperitoneally 1 h before CLP) substantially reduced mortality, especially within the first 24 h. The data shown were pooled from four experiments with wild type ($n = 40$) and C3-/- ($n = 20$), two experiments with HuC3-reconstituted C3-/- ($n = 8$) and three experiments with C4-/- mice ($n = 20$). * $P < 0.0001$ C3-/- or C4-/- versus wild type. ** $P = 0.152$ C3-/- plus HuC3 versus wild type and $P < 0.001$ C3-/- plus HuC3 versus C3-/-.

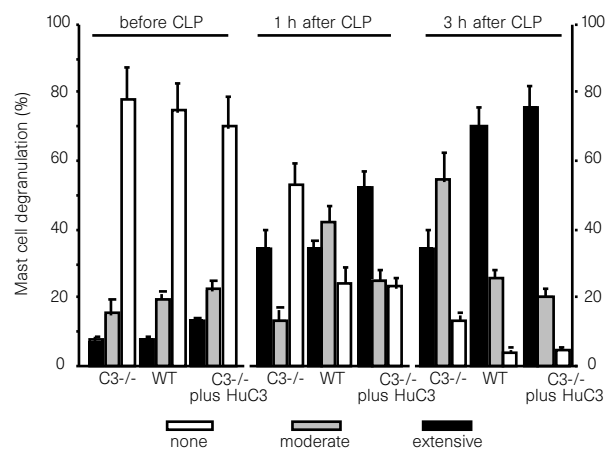


Figure 2 Activation of mast cells following CLP treatment is complement-dependent. Peritoneal lavage fluid was pooled (5 mice per group) from control mice not subjected to CLP or from mice 1 or 3 h following CLP, and the cells were concentrated by cytocentrifuge for microscopic examination after May-Grünwald-Giemsa staining. The extent of mast cell degranulation was scored as: 'none' (<10% of granules exhibiting staining alterations and/or exteriorization); 'moderate' (10-50% of granules affected); 'extensive' (>50% of granules affected) and expressed as mean $\pm$ s.e.m.¹⁹. Mast cells collected from C3-deficient mice were significantly ($P < 0.001$) less activated than those from wild-type (WT) mice; however, reconstitution with human C3 protein (as described in Fig. 1) resulted in a similar degree of mast cell activation as was observed with wild-type cells ($P > 0.05$). The data were pooled from 3-5 mice per group in each of 4 experiments with wild-type and C3-/- and 5 mice per group in each of two experiments for HuC3-reconstituted C3-/- mice, all of which gave similar results for mice in individual groups.

few or no bacteria were seen in the preparations from the wild-type or HuC3-reconstituted C3^{-/-} mice (Fig. 4a, c, d, f), the peritoneal lavage fluids ($n = 5$) from each group were pooled and cultured overnight. About 20-fold more colony-forming units (CFU) of *Escherichia coli* were identified in the peritoneal lavage fluids of C3^{-/-} mice versus wild-type mice at either 1 or 3 h following CLP (Table 1). Reconstitution of the C3-deficient mice with C3 protein not only restored infiltration of neutrophils to levels similar to those observed in the wild-type mice (Figs 3b and 4f) but also resulted in enhanced clearance of bacteria (Table 1).

Clearly, not all of the mast cell degranulation (Fig. 2) or TNF- α production (Fig. 3a) in this model is complement-dependent and this may explain, at least in part, the neutrophil recruitment that occurs when CLP is performed in C3^{-/-} mice (Fig. 3b). Multiple factors, in addition to products of the complement system, may contribute to mast cell activation and TNF- α production observed in CLP, including bacterial fimbrial adhesins¹⁸ and lipopolysaccharide (LPS)¹⁵.

Similarly, some of the important actions of complement in this model may occur independently of any effects of mast cells. Complement is thought to participate in host defence against bacterial infection at several stages, such as direct lysis of bacteria by the C5b-C9 membrane attack complex, opsonization of bacteria by C3b and iC3b ligands, and activation and chemotaxis of neutrophils through C3a and C5a peptides¹. C3-deficient mice are highly sensitive to the toxic effects of systemic endotoxin¹⁴ and to infection with group B *Streptococcus*¹¹. In addition, the serum of C3^{-/-} mice fails to support bacteriolysis by neutrophils in an antibody-dependent opsonophagocytosis assay *in vitro*¹¹. Thus, complement has some essential functions in bacterial clearance that are mast cell-independent.

C5 may have a role in the CLP model, both in assembly of the membrane attack complex on bacteria and because of the release of C5a anaphylatoxin, which can induce mast cell degranulation^{4,5} and function as an important chemo-attractant of neutrophils²³. Indeed, comparison of mice that are genetically deficient in C5 (BL10/OSN)²⁴ with controls (BL10/NSN) in the CLP model demon-

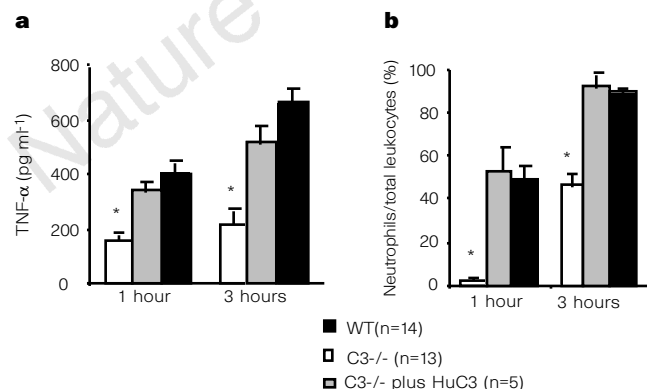


Figure 3 TNF- α release and neutrophil infiltration are impaired in C3^{-/-} mice during acute septic peritonitis. **a**, TNF- α levels in peritoneal lavage fluids were determined by ELISA and expressed as pg ml⁻¹ (2 ml total per mouse) of fluid. **b**, The ratio of neutrophils to total leukocytes within the peritoneal lavage was determined using cytocentrifuge preparations. C3-deficient mice were reconstituted with 0.5 mg human C3 protein as described in Fig. 1. Results represent mean $\pm$ s.e.m. The differences between corresponding values for C3^{-/-} plus HuC3 versus wild-type groups are not significant ($P = 0.205$). The data were pooled from four experiments with wild-type ($n = 14$) and C3^{-/-} ($n = 13$) and two experiments with HuC3-reconstituted C3^{-/-} ($n = 5$) mice, respectively, all of which gave similar results for mice in individual groups. * $P < 0.0001$ versus wild type.

Table 1 Bacterial clearance within the peritoneum is impaired in complement deficient mice

Mice	One hour after CLP	Three hours after CLP
	* CFU	* CFU
WT	1.6×10^4	5.7×10^4
C3 ^{-/-}	34.6×10^4	116×10^4
C3 ^{-/-} plus 0.5 mg HuC3	2.1×10^4	34.2×10^4

About 20-fold more colony-forming units (CFU) of *E. coli* were identified in the peritoneal lavage fluid of C3^{-/-} mice compared to wild-type mice, at either 1 or 3 h following CLP, but HuC3-reconstituted C3^{-/-} mice exhibited only ~6% or 30% of the number of CFU at 1 or 3 h after CLP than did the C3^{-/-} mice that were not treated with HuC3. CFU were determined by overnight culture of serial dilutions of peritoneal lavage fluid which had been collected from five mice in each group and then pooled before culture in one experiment. In the second experiment, CFU represents the mean of individual mice with five mice per group.

* CFU of *E. coli* after overnight culture of peritoneal lavage fluid on LB-agar plates.

strated a roughly 25-fold reduction in bacterial (*E. coli*) clearance at both 1 h (38.0×10^4 versus 1.3×10^4 CFU) and 3 h (55.6×10^4 versus 2.2×10^4 CFU), respectively, after overnight cultures of peritoneal lavage fluid pooled from five mice per group. By contrast, there was only a slight reduction in neutrophil infiltration in the C5-deficient strain relative to controls (40% versus 56% at 3 h, respectively ($P = 0.230$)). These findings, and work with C5aR^{-/-} mice²⁵, indicate that C5 can be important for efficient bacterial clearance, but may not be essential for the initial infiltration of neutrophils into the peritoneal cavity following bacterial infection.

Although the exact mechanism of complement activation in the CLP model is not yet clear, our results with C4-deficient mice indicate that the classic pathway is required (Fig. 1). It is possible that either or both of two important recognition molecules in natural immunity, natural antibody^{26,27} or mannan binding lectin²⁸ (MBL) (each of which can lead to activation of the classic pathway) may be involved in complement activation in the CLP-model. It is possible that levels of natural antibody or MBL are different in C3-deficient mice versus the mast cell-deficient and congenic normal mice which were used in previous studies of CLP¹³, but the reconstitution of C3^{-/-} mice with HuC3 largely repaired the defects in their response to CLP.

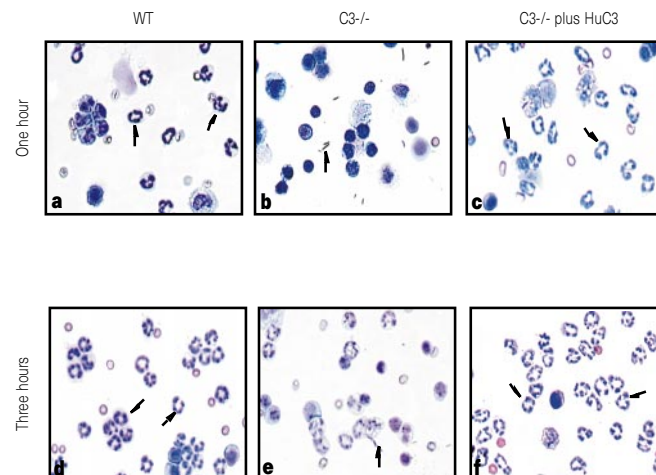


Figure 4 Neutrophil clearance of bacteria within the peritoneum is complement-dependent. Peritoneal lavage fluid was collected at 1 or 3 h and cytocentrifuge preparations were examined microscopically after staining with May-Grünwald-Giemsa. **a-f**, Representative specimens from wild-type (WT) mice (**a** and **d**); C3^{-/-} mice (**b** and **e**) or HuC3-reconstituted (as in Fig. 1) C3^{-/-} mice (**c** and **f**) taken at 1 and 3 h after CLP, respectively. Neutrophil (arrows in **a**, **c**, **d** and **f**) influx and uptake and killing of enteric bacteria (arrows in **b** and **e**) are impaired in C3^{-/-} mice; however, reconstitution of C3^{-/-} mice with human C3 protein enhances neutrophil infiltration and clearance of bacteria. Photomicrographs representative of multiple cytocentrifuge preparations as Fig. 2. Magnification, $\times 400$.

The precise mechanism of complement-dependent activation of peritoneal mast cells also remains to be determined. Although C3a and C5a induce activation of some mast cells^{4,5} including those in the mouse peritoneal cavity²¹, the C3b receptor (CD35) has also been identified on rat peritoneal mast cells²⁹ and, in preliminary studies using flow cytometry, on mouse peritoneal mast cells (Irma Gigli, personal communication). Consistent with a role for CD35 in complement-dependent mast cell activation in CLP, we have recently found that CD35-deficient (Cr2-/-)³⁰ mice are highly sensitive to CLP (95% versus 20% mortality within 24 h in Cr2-/- mice versus wild-type mice, respectively, $n = 20$ per group, $P < 0.0001$) and exhibit reduced neutrophil infiltration at 1 h or 3 h after CLP (14 or 47% of peritoneal cells in Cr2-/- versus 50 or 90% in wild-type mice at 1 or 3 h, respectively, $P < 0.003$). □

Methods

Mice. Mice deficient in C4 (C4-/-) and C3 (C3-/-) were constructed as described using the approach of homologous recombination in embryonic stem (ES) cells^{10,11}. Wild-type controls were matched for age, gender and major histocompatibility complex (MHC) and had a genetic background 129sv/C57BL/6 similar to that of the C3-/- mice. Mice genetically deficient in C5 (BL10/OSNJ)²⁴ and controls (BL10/NSNJ) were purchased from Jackson Labs (Bar Harbor, Maine). All animals were maintained in a viral-antigen-free (VAF) facility. Mice were monitored frequently for signs of morbidity or mortality. Studies were performed according to NIH and institutional guidelines for animal use and care. C3-deficient mice were reconstituted with 0.5 ml human C3 protein (HuC3 1 mg ml⁻¹) in sterile, pyrogen-free 0.9% NaCl (saline) or as a control, saline alone, 1 h before CLP. The HuC3 functional activity was estimated as >90% based on haemolytic activity, as determined by the supplier (Advanced Research Technologies, San Diego, CA).

Caecal ligation and puncture. The procedure was performed as described¹². Briefly, after inducing an appropriate level of anaesthesia with avertin (0.015 ml of a 2.5% solution per g body weight, s.c.) a 0.5 cm midline incision was made in the peritoneum and the distal two-thirds of caecum was ligated and punctured once with an 18½ gauge syringe needle. The incision was closed with a wound clip and the mice were resuscitated with 1 ml of sterile saline injected subcutaneously.

Evaluation of mast cell activation and neutrophil infiltration. Cytocentrifuge preparations of peritoneal cells were stained with May-Grünwald-Giemsa stain and examined under a light microscope at ×400 magnification to quantify the neutrophils and numbers of mast cells, and at ×1,000 to quantify the extent of mast cell degranulation¹⁹.

TNF-α assays. TNF-α levels in peritoneal lavage fluids (about 2 ml recovered from each mouse) were measured by an ELISA kit (Endogen, Cambridge, MA) according to the manufacturer's specifications.

CFU (colony-forming units). *E. coli* CFU were determined by overnight culture of serial dilutions of peritoneal lavage fluid at 37°C. Colonies of *E. coli* were identified in a background of heterogeneous colonies by morphology and confirmed by culture on MacConkey II agar and by assay with the Enteric Identification System (Organon Teknia).

Statistical analysis. Statistical analyses of most data were performed using the Student *t*-test, with two-samples and assuming unequal variances, where indicated. Statistical analysis for survival was performed using the Mantel-Cox rank test and statistical analysis of the extent of mast cell degranulation was performed using the χ^2 test¹⁹.

Received 23 June; accepted 21 August 1997.

1. Muller-Eberhard, H. J. Molecular organization and function of the complement system. *Annu. Rev. Biochem.* **57**, 321–347 (1988).
2. Reid, K. B. M. & Porter, R. R. The proteolytic activation system of complement. *Annu. Rev. Biochem.* **50**, 433–464 (1981).
3. Lepow, I. H., Wilms-Kretschmer, K., Patrick, R. A. & Rosen, F. S. Gross and ultrastructural observations on lesions produced by intradermal injection of human C3a in man. *Am. J. Pathol.* **61**, 13–23 (1970).
4. Dias Da Silva, W. & Lepow, I. H. Complement as a mediator of inflammation. II. Biological properties of anaphylatoxin prepared with purified components of human complement. *J. Exp. Med.* **125**, 921–946 (1967).
5. Johnson, A. R., Hugli, T. E. & Muller-Eberhard, H. J. Release of histamine from rat mast cells by the complement peptides C3a and C5a. *Immunology* **28**, 1067–1080 (1975).
6. Miyajima, I. *et al.* Systemic anaphylaxis in the mouse can be mediated largely through IgG1 and FcγRIII. Assessment of the cardiopulmonary changes, mast cell degranulation and death associated with active or IgG1-dependent passive anaphylaxis. *J. Clin. Invest.* **99**, 901–914 (1997).

7. Sylvestre, D. L. & Ravetch, J. V. Fc receptors initiate the Arthus reaction: redefining the inflammatory cascade. *Science* **265**, 1095–1098 (1994).
8. Hazenbos, W. L. W. *et al.* Impaired IgG-dependent anaphylaxis and arthus reaction in FcγRIII (CD16) deficient mice. *Immunology* **5**, 181–188 (1996).
9. Sylvestre, D., Clynes, R., Ma, M., Carroll, M. C. & Ravetch, J. Immunoglobulin G-mediated inflammatory responses develop normally in complement-deficient mice. *J. Exp. Med.* **184**, 2385–2392 (1996).
10. Fischer, M. B. *et al.* Regulation of the B cell response to T-dependent antigens by classical pathway complement. *J. Immunol.* **157**, 549–556 (1996).
11. Wessels, M. R. *et al.* Studies of group B streptococcal infection in mice deficient in complement C3 or C4 demonstrate an essential role for complement in both innate and acquired immunity. *Proc. Natl Acad. Sci. USA* **12**, 11490–11494 (1995).
12. Wichterman, K. A., Baue, A. E. & Chaudry, I. H. Sepsis and septic shock; a review of laboratory models and a proposal. *J. Surg. Res.* **29**, 189–201 (1980).
13. Echtenacher, B., Mannel, D. N. & Hültner, L. Critical protective role of mast cells in a model of acute septic peritonitis. *Nature* **381**, 75–77 (1996).
14. Fischer, M. B. *et al.* Increased susceptibility to endotoxin shock in complement C3 and C4 deficient mice is corrected by C1 inhibitor replacement. *J. Immunol.* **159**, 976–982 (1997).
15. Leal-Berumen, I., Conlon, P. & Marshall, J. S. Interleukin-6 production by rat peritoneal mast cells is not necessarily preceded by histamine release and can be induced by bacterial lipopolysaccharide. *J. Immunol.* **152**, 5468–5476 (1994).
16. Gordon, J. R. & Galli, S. J. Mast cells are a source of both preformed and immunologically inducible TNF-α/cachectin. *Nature* **346**, 274 (1990).
17. Gordon, J. R. & Galli, S. J. Release of both preformed and newly synthesized tumor necrosis factor α (TNF-α)/cachectin by mouse mast cells stimulated by the FcεRI. A mechanism for the sustained action of mast cell-derived TNF-α during IgE-dependent biological responses. *J. Exp. Med.* **174**, 103–107 (1991).
18. Malaviya, R., Ikeda, T., Ross, E. & Abraham, N. Mast cell modulation of neutrophil influx and bacterial clearance at sites of infection through TNF-α. *Nature* **381**, 77–80 (1996).
19. Wershil, B. K., Wang, Z.-S., Gordon, J. R. & Galli, S. J. Recruitment of neutrophils during IgE-dependent cutaneous late phase responses in the mouse is mast cell-dependent: partial inhibition of the reaction with antiserum against tumor necrosis factor-α. *J. Clin. Invest.* **87**, 446–453 (1991).
20. Zhang, Y., Ramos, B. F. & Jakschik, B. A. Augmentation of reverse passive arthus reaction by mast cells in mice. *J. Clin. Invest.* **88**, 841–846 (1991).
21. Ramos, B. F., Zhang, Y. & Jakschik, B. A. Neutrophil elicitation in the reverse passive arthus reaction. *J. Immunol.* **152**, 1380–1384 (1994).
22. Vassalli, P. The pathophysiology of tumor necrosis factor. *Annu. Rev. Immunol.* **10**, 411–452 (1992).
23. Gerard, C. & Gerard, N. P. 5a anaphylatoxin and its seven transmembrane-segment receptor. *Annu. Rev. Immunol.* **12**, 775–808 (1994).
24. Nilsson, U. R. & Muller-Eberhard, H. J. Deficiency of the fifth component of complement in mice with an inherited complement defect. *J. Exp. Med.* **125**, 1–16 (1967).
25. Hopken, U. E., Lu, B., Gerard, N. P. & Gerard, C. The C5a chemoattractant receptor mediates defence to infection. *Nature* **383**, 86–89 (1996).
26. Briles, D. *et al.* Antiphosphocholine antibodies found in normal mouse serum are protective against intravenous infection with type 3 *Streptococcus pneumoniae*. *J. Exp. Med.* **153**, 694–705 (1981).
27. Reid, R. R. *et al.* Unraveling the role of natural antibody and complement in the clearance of lipopolysaccharide. *J. Immunol.* **159**, 970–975 (1997).
28. Epstein, J., Eichbaum, Q. E., Sherif, S. & Ezekowitz, R. A. B. The collectins in innate immunity. *Curr. Opin. Immunol.* **8**, 29–35 (1996).
29. Vranian, G., Conrad, D. H. & Ruddy, S. Specificity of C3 receptors that mediate phagocytosis by rat peritoneal mast cells. *J. Immunol.* **126**, 2302–2306 (1981).
30. Ahearn, J. *et al.* Disruption of the Cr2 locus results in a reduction in B-1a cells and in an impaired B cell response to T-dependent antigen. *Immunology* **4**, 251–262 (1996).

Acknowledgements. We thank M. Ma, J. Xia and L. Fox for excellent technical assistance. This study was supported by grants from the National Institutes of Health and Cambridge Antibodies Technology, UK.

Correspondence and requests for materials should be addressed to M.C.C. (e-mail: mcarroll@warren.med.harvard.edu).

A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function

Dirk M. Anderson, Eugene Maraskovsky*, William L. Billingsley, William C. Dougall, Mark E. Tometsko, Eileen R. Roux*, Mark C. Teepe*, Robert F. DuBose†, David Cosman & Laurent Galibert

Departments of Molecular Biology, *Immunobiology and †Bioinformatics, Immunex Corporation, 51 University St, Seattle, Washington 98101, USA

Dendritic cells are rare haematopoietic cells that reside in a number of organs and tissues. By capturing, processing and presenting antigens to T cells, dendritic cells are essential for immune surveillance and the regulation of specific immunity^{1–4}. Several members of the tumour necrosis factor receptor (TNFR) superfamily are integral to the regulation of the immune response. These structurally related proteins modulate cellular functions ranging from proliferation and differentiation to

The unique capacity of dendritic cells to initiate or regulate immunity has led to their study as sources of genes involved in the modulation of the immune response. Because only extremely small numbers of dendritic cells can be procured from normal individuals or *in vitro* culture regimens, traditional methods of gene discovery requiring large numbers of cells are not practical. We therefore used direct sequencing of a dendritic-cell cDNA library to identify new genes. While sequencing cDNAs from a human bone-marrow-derived myeloid dendritic-cell cDNA library, we identified a partial cDNA encoding an open reading frame (ORF) with homology to a portion of the extracellular domain of human CD40, a member of the TNFR superfamily. The full-length RANK cDNA encodes a type I transmembrane protein of 616 amino acids with a predicted extracellular domain of 184 amino acids and a cytoplasmic domain of 383 amino acids (Fig. 1a). RANK bears significant amino-acid homology with other members of the TNFR family in its four extracellular cysteine-rich pseudorepeats⁵. When the hRANK extracellular domain was compared with those of other TNFR family members, amino-acid identities ranged from 40% to 21%, with highest identity to CD40 and lowest to CD27 (data not

To identify the ligand for RANK (RANKL), we first tested whether a recombinant soluble version of RANK (RANK.Fc) could bind other members of the TNF ligand family (cells transfected with cDNAs encoding transmembrane versions of Fas ligand (FasL), CD27L, CD30L, CD40L, 4-1BBL, OX40L, TNF- α , LT- $\alpha + \beta$, or TRAIL). No binding was detected. By using flow cytometry with RANK.Fc and biotinylated anti-human IgG1 antibody, we detected a high level of RANK.Fc binding on the murine thymoma line EL40.5 (ref. 9 and data not shown). We screened an EL40.5 expression cDNA library using a two-step slide transfection and detection method, and isolated a cDNA that directed the expression of a cell-surface molecule capable of binding RANK.Fc. The full-length murine RANKL (mRANKL) cDNA encodes a type II transmembrane protein of 316 amino acids with a predicted cytoplasmic domain of 48 amino acids and an extracellular domain of 247 amino acids, and shares 85% identity with its human counterpart (Fig. 2a). When the hRANKL extracellular domain was compared with other TNF ligand-family extracellular domains, amino-acid identities ranged from 28% to 18%, with highest identity to CD40L and lowest to CD27L (data not shown). RANKL mRNA was detected in lymphoid cell lines and tissues on a panel of murine cell-line and primary cell mRNAs (Fig. 2b). Among

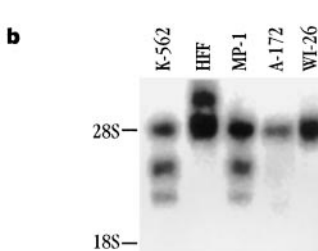


Figure 1 Sequence and expression of RANK. **a**, Amino-acid sequence alignment of hRANK and mRANK. A murine homologue of the hRANK cDNA was isolated by PCR amplification and colony hybridization from a murine fetal liver epithelium cDNA library. The predicted amino-acid sequences of hRANK and mRANK are 70% identical. The four predicted cysteine-rich pseudorepeats (labelled I-IV) are indicated, with conserved cysteine residues indicated. Predicted signal peptide and membrane-spanning regions are shown, and potential N-linked glycosylation sites are shaded. The RANK cytoplasmic domain bears little sequence similarity to other TNFR family members and has no homology with other known proteins. **b**, Analysis of hRANK mRNA expression. Poly(A⁺) mRNA (1 µg per lane) was blotted and probed with a radiolabelled RANK RNA probe. The relative migration of 18S and 28S rRNA is indicated. K-562, chronic myelogenous leukaemia (ATCC CCL 243); HFF, human foreskin fibroblasts; MP-1, B lymphoblastoid line²⁶; A-172, glioblastoma (ATCC CRL 1620); WI-26, SV40-transformed lung fibroblast (ATCC CCL 95.1).

human primary tissue mRNAs, a strong RANKL-specific signal was seen in the lymph node, with weak hybridization detectable in spleen, peripheral blood leukocytes, bone marrow, heart, placenta, skeletal muscle, stomach and thyroid (data not shown). Thus RANKL mRNA seems to be more restricted than RANK mRNA, with expression primarily in T cells, T-cell lines and lymphoid tissues. The chromosomal locations of RANK and RANKL were mapped using a human/hamster somatic cell panel and radiation hybrid mapping panel. RANK mapped to human chromosome 18q22.1; RANKL mapped to 13q14. These are the first TNFR and TNF-ligand family member genes mapped to these regions.

Many of the biological effects exerted by TNFR family members are mediated through the transcription factor NF- κ B^{10,11}. In an electrophoretic mobility-shift assay, RANK overexpression was sufficient to activate NF- κ B, and ligand-dependent NF- κ B activation was demonstrated by co-transfection of increasing concentrations of transmembrane RANKL with RANK (Fig. 3a). NF- κ B/DNA complexes were also induced in RANK⁺ peripheral blood T lymphocytes (PBTs) incubated with soluble human RANKL (hRANKL) (Fig. 3a), indicating RANK-specific activation of NF- κ B in human T cells.

To identify cells responsive to RANKL, a panel of 26 cell lines and primary cells representing several cell types were screened for surface RANK expression. Only dendritic cells, VK1 (human CD4⁺ T cell line), MP-1 and human foreskin fibroblasts expressed RANK protein on their surface (data not shown). We could not detect RANK protein on two cell lines (K-562 and WI-26 V4A) that expressed a considerable amount of RANK mRNA (Fig. 1b), suggesting that surface RANK expression may be post-transcriptionally regulated. Because many TNFR family proteins are modulated during cellular activation¹², we examined whether RANK surface expression could be upregulated on freshly isolated PBTs *in vitro*. PBTs stimulated at low cell density for 7 days in the presence of phytohaemagglutinin (PHA, 1% w/v) or on plates coated with anti-CD3 with or without interleukin (IL)-1 α , -2, -3, -4, -5, -6, -7, -8, -10, -12, -13, -15, TNF- α , interferon- γ or TGF- β 1 were examined for RANK surface expression by flow cytometry. We did not detect RANK surface expression in the absence of cytokines. Of the cytokines tested, only IL-4 and TGF- β consistently induced surface expression of RANK on either PHA- or anti-CD3-activated

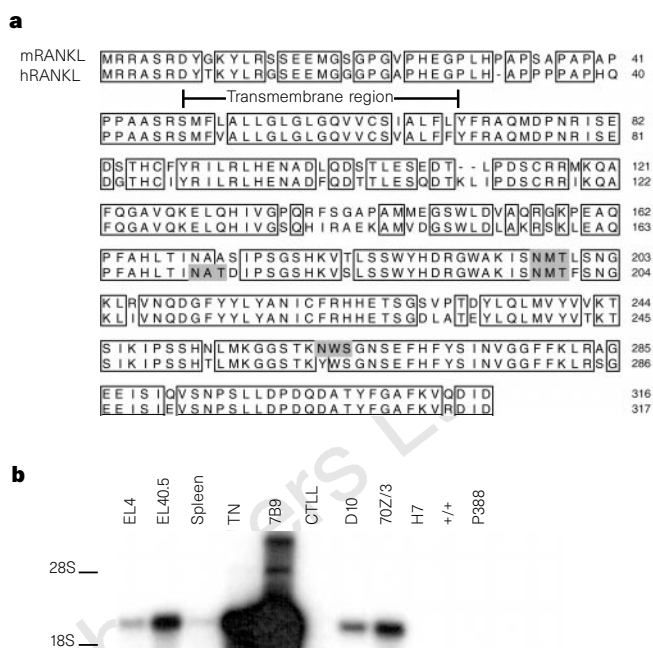


Figure 2 Sequence and expression of RANKL. **a**, Amino-acid sequence alignment of mRANKL and hRANKL. A human homologue of the mRANKL cDNA was isolated by PCR amplification and plaque hybridization from a human PBL cDNA library. The full-length hRANKL cDNA shares 85% amino-acid identity with mRANKL. The predicted membrane-spanning region is indicated, and predicted N-linked glycosylation sites are shaded. **b**, mRANKL mRNA expression. EL4, lymphoma (ATCC TIB 39); EL4.5, EL4 sorted for high CD40L expression⁸; TN, [Thy 1.2, CD25, CD44]⁺, [CD3, CD4, CD8]⁺ T progenitors²²; 7B9, T helper clone²⁸; CTLL, cytolytic T lymphocyte clone (ATCC TIB 214); D10, T helper clone (D10.G4.1, ATCC TIB 224); 70Z/3, pre-B cell line (ATCC TIB 158); H7, murine mast cell line; +/+, bone marrow stromal cell line²²; P388, macrophage line, (P388D₁, ATCC TIB 63).

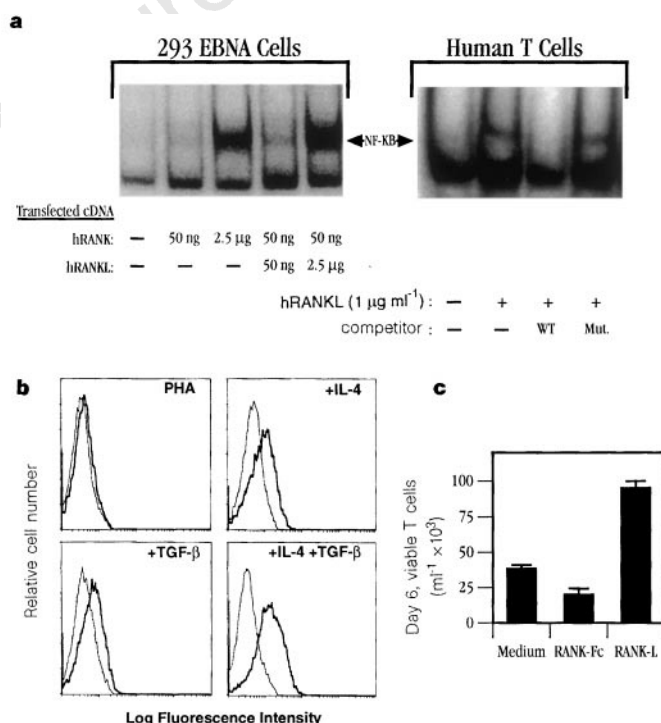


Figure 3 Functional expression of RANK in transfected 293EBNA and in PBTs.

a, RANK/RANKL-mediated NF- κ B activation. Electrophoretic mobility shift assays (EMSA)²⁹ were performed on 293EBNA cells 48 h after transfection with the indicated cDNAs. The specificity of the reaction was confirmed by competition with 50-fold excess of unlabelled oligonucleotides containing wild-type or mutated NF- κ B binding sites (not shown). PBTs treated for 5 days with 1% PHA and 10 ng ml⁻¹ IL-4 were assayed by EMSA after no treatment or treatment with soluble hRANKL for 30 min. Nuclear extracts were assayed in the absence or presence of unlabelled wild-type (WT) or mutant (Mut.) competitor oligonucleotides. **b**, Induction of RANK on PBTs. Human PBTs were cultured for 6 days at 2×10^5 cells ml⁻¹ in the presence of 1% (w/v) PHA, with or without cytokines. The expression of RANK was then assessed by flow cytometry using biotinylated anti-hRANK monoclonal antibody (bold lines) or unrelated isotype-matched control antibodies (thin lines) followed by streptavidin-R phycoerythrin. **c**, Influence of RANK.Fc and hRANKL on activated T-cell growth. PBTs were cultured at 7×10^5 cells per well for 6 days in 24-well plates precoated with anti-CD3 (OKT3, 5 μ g ml⁻¹) and anti-FLAG (M1, 5 μ g ml⁻¹; for crosslinking recombinant RANKL) in the presence of TGF- β (1 ng ml⁻¹) and IL-4 (10 ng ml⁻¹), with or without recombinant FLAG-tagged soluble hRANKL (1 μ g ml⁻¹) or RANK.Fc (10 μ g ml⁻¹). Viable T-cell recovery was determined by triplicate trypan blue countings.

PBTs (Fig. 3b). When combined, these two cytokines displayed a synergistic effect. RANK was detectable as early as day 4 and peaked by day 6 (not shown). The addition of TGF- β to anti-CD3- or PHA-activated human PBTs has been shown to induce arrest of proliferation and ultimately death of most lymphocytes within the first few days of culture.¹³ We therefore examined the effect of interactions between RANK and RANKL on TGF β -treated PBTs by adding RANK.Fc or soluble hRANKL to PBT cultures. The addition of RANK.Fc significantly reduced the number of viable T cells recovered after 6 days, whereas soluble RANKL greatly increased the viable cell recovery (Fig. 3c). Thus RANK activation enhances the number of viable T cells generated in the presence of TGF- β . It remains to be established whether this is a survival or a proliferative effect. IL-4 and TGF- β , which have the unique capacity to induce RANK expression on activated T cells, have been implicated in the regulation of immune response when secreted by the TH3/regulatory T-cell subset. These T cells are believed to mediate bystander suppression of effector T cells^{14,15}. Accordingly, our observations suggest that the interaction between RANK and RANKL could be implicated in the induction of T-cell tolerance.

On human CD1a⁺ dendritic cells derived from cultured CD34⁺ bone-marrow cells, only a subset (20–30%) of CD1a⁺ dendritic cells expressed RANK at the cell surface. However, the addition of CD40L to the dendritic-cell cultures greatly increased RANK surface expression (Fig. 4a). CD40L activates dendritic cells by increasing cluster formation *in vitro*, inducing morphological changes to dendritic cells, and upregulating expression of HLA-DR, CD54, CD58, CD80 and CD86, all of which seem to be directly correlated with an increased capacity of dendritic cells to stimulate the proliferation of alloreactive T cells^{7,8}. This effect upon the function

of dendritic cells has been reported only for CD40L^{7,8}. In comparison, soluble hRANKL significantly increased the degree of dendritic-cell aggregation and cluster formation above control cultures, comparable to CD40L cultures (Fig. 4b). This effect was not evident when either heat-inactivated RANKL (Fig. 4b) or RANK.Fc antagonist (data not shown) were used, confirming that it was dependent on biologically active protein. However, unlike that induced by CD40L, RANKL-induced clustering was not due to increased expression of CD2, CD11a, CD54 or CD58 (data not shown). The addition of RANKL to CD1a⁺ dendritic cells increased their allostimulatory capacity in a mixed lymphocyte reaction by between three- and tenfold, comparable to CD40L-cultured dendritic cells (Fig. 4c). The allostimulatory capacity of dendritic cells was not increased when they were cultured with either heat-inactivated RANKL (Fig. 4c) or with RANKL plus RANK.Fc antagonist (data not shown). The allostimulatory activity of dendritic cells was not further increased when RANKL and CD40L were used in combination, possibly because the functional capacity of dendritic cells reached a maximal level with either cytokine alone. Neither RANKL or CD40L significantly enhanced the *in vitro* growth of dendritic cells over the culture period of 3 days before addition to the mixed lymphocyte reaction (data not shown). Unlike CD40L, RANKL did not significantly increase the level of expression of HLA-DR, CD80 or CD86 (data not shown). It is intriguing that RANKL can enhance dendritic-cell cluster formation and functional capacity without modulating known molecules involved in cell adhesion (CD18 and CD54), antigen presentation (HLA-DR) or co-stimulation (CD80 and cD86), all of which are regulated by CD40/CD40L signalling^{7,8}.

The molecular basis for the unique ability of dendritic cells to

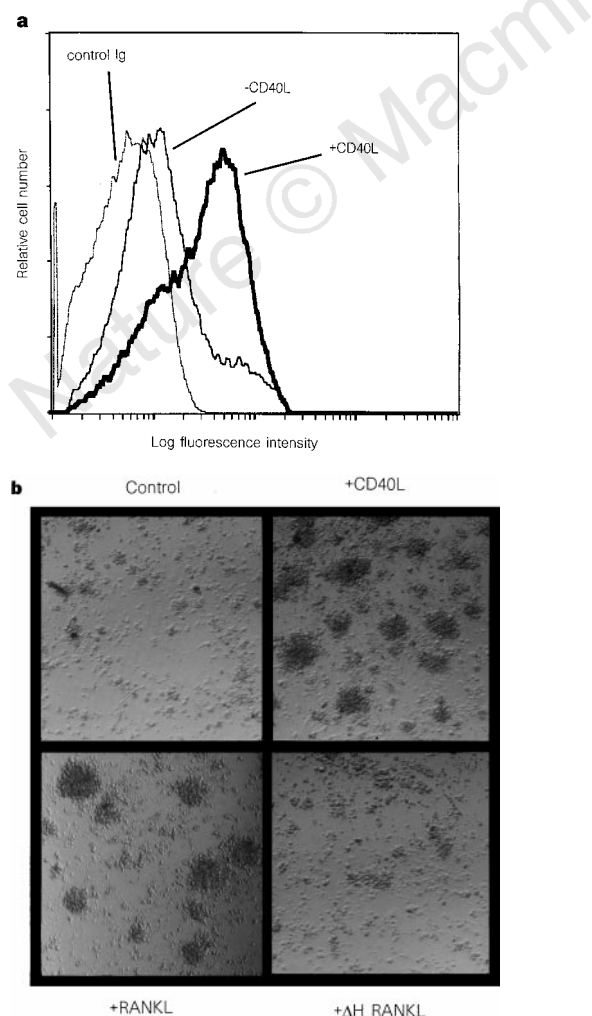


Figure 4 RANK expression and function on CD1a⁺ dendritic cells. **a**, CD40L upregulates RANK expression on human dendritic cells. Sorted human CD1a⁺ dendritic cells derived from CD34⁺ bone-marrow cells cultured in GM-CSF, IL-4, TNF- α and FL for 14 days were examined for RANK surface expression in the presence or absence of CD40L for 48 h, by flow cytometry using a biotinylated anti-RANK monoclonal antibody followed by streptavidin-R phycoerythrin. A biotinylated IgG isotype-matched control monoclonal antibody was used as a background control. **b**, RANKL induces human dendritic-cell cluster formation. Sorted human CD1a⁺ dendritic cells were cultured in a cytokine cocktail (GM-CSF, IL-4, TNF- α and FL) (top left), in cocktail plus CD40L (1 μ g ml⁻¹) (top right), in cocktail plus RANKL (1 μ g ml⁻¹) (bottom left), or in cocktail plus heat-inactivated (Δ H) RANKL (1 μ g ml⁻¹) (bottom right) in 24-well flat-bottomed culture plates in 1 ml culture media for 48–72 h and then photographed using an inversion microscope. **c**, RANKL increases the allostimulatory capacity of dendritic cells. Dendritic cells were cultured as in **b** and then washed, irradiated (2,000 rad), and incubated with allogeneic T cells (1 $\times 10^5$ cells per well) in increasing numbers in 96-well round-bottomed culture plates in 0.2 ml culture medium for 4 days. The cultures were pulsed with 0.5 μ Ci [³H]-thymidine for 8 h, collected and gas-phase-beta counted. The background counts for either T cells or dendritic cells cultured alone were <100 c.p.m. Values represent the mean $\pm$ s.d. of triplicate cultures, and are representative of experiments performed with dendritic cells derived from 7 different bone-marrow donors.

prime naive T cells is only partly understood, but it is commonly attributed to the fact that dendritic cells can express high levels of co-stimulatory molecules when interacting with CD40L⁺ antigen-specific T cells. Our results not only define a new TNFR member expressed on dendritic cells, but also provide evidence for a molecular interaction between dendritic cells and T cells that appears to regulate CD80/86-independent T-cell priming. Given that CD40L regulates RANK surface expression on *in vitro*-generated dendritic cells and that CD40L is upregulated on activated T cells during interactions between dendritic cells and T cells, RANK and its ligand may form an important part of the activation cascade that is induced during dendritic cell-mediated T-cell expansion. Further work is required before we understand fully the role of RANK and its ligand and their biological effects upon dendritic cells and T cells.

Note added in proof: A molecule with the same sequence as RANKL has recently been described³⁰. □

Methods

Cell culture, antibodies and growth factors. Mature human T cells were purified from peripheral blood or bone marrow of healthy donors by E-rosetting on sheep red blood cells¹⁶ followed by depletion with anti-CD14, -CD19 and -CD16 monoclonal antibodies (Immunotech) coupled to magnetic beads (Dyna). T cells were cultured at 2×10^5 cells ml⁻¹ either in the presence of 1% PHA or on anti-CD3 (OKT3, 5 µg ml⁻¹) pre-coated plates. Cytokines were produced at Immunex, with the exception of IL-2 (Cetus), IL-4, IL-8, IL-10 and TGF-β1 (Genzyme), IL-12 (R&D Systems), and interferon-γ (Sigma).

Isolation of human dendritic cells. Functionally mature dendritic cells were generated from CD34⁺ bone-marrow progenitors. Human bone-marrow cells from healthy volunteers were density fractionated using Ficoll medium, and CD34⁺ immunoaffinity isolated using an anti-CD34 matrix column (Cephar, CellPro). The CD34⁺ cells were then cultured in GM-CSF (20 ng ml⁻¹), IL-4 (20 ng ml⁻¹), TNF-α (20 ng ml⁻¹), CHO-derived Flt3L (100 ng ml⁻¹) and c-kit ligand (100 ng ml⁻¹) in Super McCoy's medium supplemented with 10% fetal calf serum in a fully humidified 37 °C incubator (5% CO₂) for 14 days. CD1a⁺, HLA-DR⁺ dendritic cells were then sorted using a FACStar Plus, and used for the construction of dendritic-cell cDNA libraries and for biological evaluation of RANK and RANKL.

cDNA library construction, sequence analysis, and RNA analysis. Poly(A)⁺ dendritic-cell RNA was converted to double-stranded cDNA with random hexamer primers and reverse transcriptase, modified with adaptors¹⁷ and ligated to BamHI-linearized and adapted pBluescript (Stratagene). The recombinants were transformed into *Escherichia coli*, and miniprep plasmid DNA from individual colonies was prepared for sequencing. Sequence analysis was performed as described¹⁸. Dendritic-cell cDNA single-pass sequences were searched against a non-redundant version of the GenBank nucleotide sequence database, dbEST and non-redundant protein databases using the BLAST algorithm.¹⁹ The murine thymoma subline EL40.5 expression cDNA library⁹, the human peripheral blood lymphocyte (PBL) cDNA library²⁰ and the WI-26 cDNA library²¹ have been described. Polyadenylated RNAs were prepared, electrophoresed, blotted and probed as described¹⁸.

Recombinant protein production and purification. A RANK.Fc fusion protein was used for antibody production and cell-binding experiments. A fragment encoding the predicted extracellular domain of hRANK (amino acids 1–213) was generated by the polymerase chain reaction (PCR) and fused to the Fc coding domain of human IgG1 in mammalian expression vector pDC409 (ref. 22). RANK.Fc fusion protein was purified from pDC409-hRANK.Fc-transfected CV-1/EBNA²³ cell supernatants by protein A affinity chromatography. To generate recombinant soluble hRANKL, the extracellular coding domain (amino acids 136–317 of hRANKL, encoding the region of conserved secondary structure of the TNF ligand family²⁴) was expressed with an N-terminal FLAG peptide for detection and crosslinking and a six-histidine tag for ease of purification.²⁵ Transfected cell supernatants containing secreted FLAG-(His)₆-hRANKL were purified by chromatography using Ni-NTA Superflow bulk resin (Qiagen).

Expression library screening. The EL40.5 cDNA library was screened by a slide transfection and two-step detection method using RANK.Fc and [¹²⁵I]-

goat anti-human IgG1 antibody, essentially as described²⁶. cDNA pools positive for RANK.Fc binding were divided and rescreened until a single positive clone was isolated.

Chromosomal localization. Initial human chromosomal assignments for RANK and RANKL were made using hRANK- or hRANKL-specific PCR primers and a BIOS somatic cell hybrid PCRable DNA kit (BIOS Laboratories), following the manufacturer's instructions. More detailed mapping was performed using the Genebridge 4 radiation hybrid panel²⁷ (Research Genetics). Data from this analysis was then submitted electronically to the MIT radiation hybrid mapper (<http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl>) following the instructions contained therein. This analysis yielded specific genetic marker names which, when submitted electronically to the NCBI Entrez browser (<http://www3.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=c&form=0>), yielded the specific map locations.

Received 19 August; accepted 14 October 1997.

- Steinmann, R. M. The dendritic cell system and its role in immunogenicity. *Annu. Rev. Immunol.* **9**, 271–296 (1991).
- Caux, C. & Banchereau, J. in *Blood Cell Biochemistry* Vol. 7, *Hemopoietic growth factors and their receptors* (eds Whetton, A. & Gordon, J.) 263 (Plenum, London, 1996).
- Dubois, B. *et al.* Dendritic cells enhance growth and differentiation of CD40-activated B lymphocytes. *J. Exp. Med.* **185**, 941–951 (1997).
- Austyn, J. M. New insights into the mobilization and phagocytic activity of dendritic cells. *J. Exp. Med.* **183**, 1287–1292 (1996).
- Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* **76**, 959–962 (1994).
- Baker, S. J. & Reddy, E. P. Transducers of life and death: TNF superfamily and associated proteins. *Oncogene* **12**, 1–9 (1996).
- Caux, C. *et al.* Activation of human dendritic cells through CD40 cross-linking. *J. Exp. Med.* **180**, 1263–1272 (1994).
- Cella, M. *et al.* Ligation of CD40 on dendritic cells triggers production of high levels of interleukin-12 and enhances T cell stimulatory capacity: T-T help via APC activation. *J. Exp. Med.* **184**, 747–752 (1996).
- Armitage, R. J. *et al.* Molecular and biological characterization of a murine ligand for CD40. *Nature* **357**, 80–82 (1992).
- Baeuerle, P. A. & Henkel, T. Function and activation of NF-κB in the immune system. *Annu. Rev. Immunol.* **12**, 141–179 (1994).
- Rothe, M., Sarma, V., Dixit, V. M. & Goeddel, D. V. TRAF2-mediated activation of NF-κB by TNF receptor 2 and CD40. *Science* **269**, 1424–1427 (1995).
- Armitage, R. J. Tumor necrosis factor receptor superfamily members and their ligands. *Curr. Opin. Immunol.* **6**, 407–413 (1994).
- de Jong, R. *et al.* Differential effect of transforming growth factor-β1 on the activation of human naive and memory CD4⁺ T lymphocytes. *Int. Immunol.* **6**, 631–638 (1994).
- Fukaura, H. *et al.* Induction of circulating myelin basic protein and proteolipid protein-specific transforming growth factor-β-secreting Th3 T cells by oral administration of myelin in multiple sclerosis patients. *J. Clin. Invest.* **98**, 70–77 (1996).
- Weiner, H. L. *et al.* Oral tolerance: immunologic mechanisms and treatment of animal and human organ-specific autoimmune diseases by oral administration of autoantigens. *Annu. Rev. Immunol.* **12**, 809–837 (1994).
- Madsen, M., Johnsen, H. E., Hansen, P. W. & Christiansen, S. E. Isolation of human T and B lymphocytes by E-rosette gradient centrifugation. Characterization of the isolated subpopulations. *J. Immunol. Methods* **33**, 323–336 (1980).
- Larsen, A. *et al.* Expression cloning of a human granulocyte colony-stimulating factor receptor: a structural mosaic of hematopoietin receptor, immunoglobulin, and fibronectin domains. *J. Exp. Med.* **172**, 1559–1570 (1990).
- Anderson, D. M. *et al.* Functional characterization of the human interleukin-15 receptor α chain and close linkage of IL15RA and IL2RA genes. *J. Biol. Chem.* **270**, 29862–29869 (1995).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- Idzerda, R. L. *et al.* Human interleukin 4 receptor confers biological responsiveness and defines a novel receptor superfamily. *J. Exp. Med.* **171**, 861–873 (1990).
- Goodwin, R. G. *et al.* Cloning of the human and murine interleukin-7 receptors: demonstration of a soluble form and homology to a new receptor superfamily. *Cell* **60**, 941–951 (1990).
- Giri, J. G. *et al.* Utilization of the β and γ chains of the IL-2 receptor by the novel cytokine IL-15. *EMBO J.* **13**, 2822–2830 (1994).
- McMahan, C. J. *et al.* A novel IL-1 receptor, cloned from B cells by mammalian expression, is expressed in many cell types. *EMBO J.* **10**, 2821–2832 (1991).
- Baum, P. R. *et al.* Molecular characterization of murine and human OX40/OX40 ligand systems: identification of a human OX40 ligand as the HTLV-1 regulated protein gp34. *EMBO J.* **13**, 3992–4001 (1994).
- Hoffmann, A. & Roeder, R. G. Purification of his-tagged proteins in non-denaturing conditions suggests a convenient method for protein interaction studies. *Nucleic Acids Res.* **19**, 6337–6338 (1991).
- Goodwin, R. G. *et al.* Molecular and biological characterization of a ligand for CD27 defines a new family of cytokines with homology to tumor necrosis factor. *Cell* **73**, 447–456 (1993).
- Walter, M. A., Spillet, D. J., Thomas, P., Weisenbach, J. & Goodfellow, P. N. A method for constructing radiation hybrid maps of whole genomes. *Nature Genet.* **7**, 22–28 (1994).
- Mosley, B. *et al.* The murine interleukin-4 receptor: molecular cloning and characterization of secreted and membrane bound forms. *Cell* **59**, 335–348 (1989).
- Yao, Z. *et al.* Herpesvirus saimiri encodes a new cytokine, IL-17, which binds to a novel cytokine receptor. *Immunity* **3**, 811–821 (1995).
- Wong, B. R. *et al.* TRANCE is a novel ligand of the tumor necrosis factor receptor family that activates c-Jun N-terminal kinase in T cells. *J. Biol. Chem.* **272**, 25190–25194 (1997).

Acknowledgements. We thank L. T. Hollingsworth, C.-P. Huang, M. S. Timour and J. S. Bertles for DNA sequencing; S. C. Braddy and D. E. Hirschstein for cell sorting; M. J. Petersen, S. A. Sherer, Z. Sadeghi and P. J. Smolak for technical assistance; M. Hall for graphical assistance; A. Bannister for editorial assistance; and M. Caligiuri of the Roswell Park Cancer Institute for human bone-marrow samples.

Correspondence and requests for materials should be addressed to D.M.A. (e-mail: andersond@immunex.com). Sequences have been deposited in GenBank under the accession numbers AF018253, AF019046, AF019047 and AF019048.

DAP kinase links the control of apoptosis to metastasis

Boaz Inbal*, Ofer Cohen*, Sylvie Polak-Charcon†, Juri Koplovic†, Ezra Vadai‡, Lea Eisenbach‡ & Adi Kimchi*

* Departments of Molecular Genetics and ‡ Immunology, Weizmann Institute of Science, Rehovot 76100, Israel

† Department of Histopathology, Chaim Sheba Medical Center, Tel-Hashomer, Israel

DAP kinase is a new type of calcium/calmodulin-dependent enzyme that phosphorylates serine/threonine residues on proteins. Its structure contains ankyrin repeats and the 'death' domain, and it is associated with the cell cytoskeleton¹⁻³. The gene encoding DAP kinase was initially isolated as a positive mediator of apoptosis induced by interferon- γ , by using a strategy of functional cloning⁴. We have now tested whether this gene has tumour-suppressive activity. We found that lung carcinoma clones, characterized by their highly aggressive metastatic behaviour and originating from two independent murine lung tumours, did not express DAP kinase, in contrast to their low-metastatic counterparts. Restoration of DAP kinase to physiological levels in high-metastatic Lewis carcinoma cells suppressed their ability to form lung metastases after intravenous injection into syngeneic mice, and delayed local tumour growth in a foreign 'microenvironment'. Conversely, *in vivo* selection of rare lung lesions following injection into syngeneic mice of low-metastatic Lewis carcinoma cells or of DAP kinase transfectants, was associated with loss of DAP kinase expression. *In situ* TUNEL staining of tumour sections revealed that DAP kinase expression from the transgene raised the incidence of apoptosis *in vivo*. DAP-kinase transfectants also showed increased sensitivity *in vitro* to apoptotic stimuli, of the sort encountered by metastasizing cells at different stages of malignancy. We propose that loss of DAP kinase expression provides a unique mechanism that links suppression of apoptosis to metastasis.

Some tumour-suppressor genes (for example, those encoding p53 or APC) and oncogenes (for example, members of the Bcl-2 family) mediate or suppress apoptosis, respectively⁵⁻¹³, and thus establish a link between tumorigenesis and loss of apoptotic responses. We investigated whether genes that were initially identified as positive mediators of apoptosis could have tumour-suppressive activity. One such gene, encoding DAP kinase¹⁻³, was studied because its expression is frequently lost in human carcinoma and B-cell leukaemia cell lines¹⁴. We analysed two independent sets of high- and low-metastatic clones, selected from the murine Lewis (3LL)¹⁵ and CMT64¹⁶ lung carcinoma cell lines, and an intriguing pattern emerged: DAP kinase protein was undetectable in the two high-metastatic derivatives, whereas the low-metastatic counterparts were positive for DAP kinase (Fig. 1a). No DAP kinase messenger RNA could be detected in the high-metastatic clones (not shown). These initial observations prompted us to test whether the introduction of a functional DAP kinase gene into one of the high-metastatic clones might reduce its tumorigenic and metastatic capacity.

The Flag-tagged wild-type DAP kinase gene was transfected into the 3LL clone-D122 cells. Stable DAP-kinase-positive clones, which displayed normal growth kinetics in culture, were isolated (Fig. 1b). The transgene was expressed at physiological levels, as evaluated by comparison with the endogenous DAP kinase protein levels in the A9-F low-metastatic clone. Expression ranged between 0.7 (clone 6-DAPk), 1.5 (clone 48-DAPk), 1.8–3 (28-DAPk) and 3–6 (42-DAPk) times endogenous levels (calculated from Fig. 1b and other immu-

noblots; results not shown). *In vitro* kinase assays² proved that the DAP kinase protein, expressed from the transgene, was catalytically active (not shown). The ectopic expression of DAP kinase in clone 42-DAPk did not evoke any signs of chromatin condensation or fragmentation (Fig. 3c, bottom left panel), indicating that DAP kinase by itself, even in the clone that displayed the highest expression levels, did not trigger apoptosis under normal growth conditions.

C57BL/6 syngeneic mice were then injected with the transfected D122 cells using two different protocols¹⁵. One group of mice received intravenous (i.v.) injections and experimental metastases in the lungs were monitored. A second group received intrafootpad (i.f.p.) injections to investigate local tumour growth in the foreign microenvironment of footpad muscles, and to monitor post-surgical spontaneous metastasis in the lungs. The experiments were repeated three times with reproducible results.

In the protocol measuring experimental metastasis, the lungs were examined 30–32 days after i.v. injection. Whereas the parental D122- and G418-resistant control clones were highly metastatic, the metastatic activity of DAP kinase transfectants was strongly suppressed (Fig. 2a, b). Both the average lung weight (Fig. 2a) and the number of lung lesions (Table 1) were reduced. The extent of suppression was proportional to the levels of exogenously expressed DAP kinase, and even the low-expressor clone 6-DAPk had a significant reduction in its metastatic activity (Table 1). Mice injected with the 42-DAPk clone remained viable for at least 112 days post-injection; their lungs at this time contained only 0–2 nodules.

In the second protocol, which consisted of intrafootpad injections, the effect of DAP kinase was milder, being restricted to clones that expressed the highest DAP kinase levels (28- and 42-DAPk). The effect was manifested by a delay in growth of local tumours, which was directly proportional to the levels of exogenously expressed DAP kinase (Fig. 2c and inset). Once tumours emerged, however, the kinetics of increase in total tumour mass were similar to those observed in the control clones. Moreover, surgical removal of the late-appearing tumours, generated by clones 28- or 42-DAPk, gave rise to spontaneous lung metastases which eventually killed the mice. These observations further suggested that revertants, in which the transgene was lost or inactivated, might have been selected *in vivo* after i.f.p. injection. Indeed, examination of tumour cells, recovered in culture from the post-surgical lung metastases,

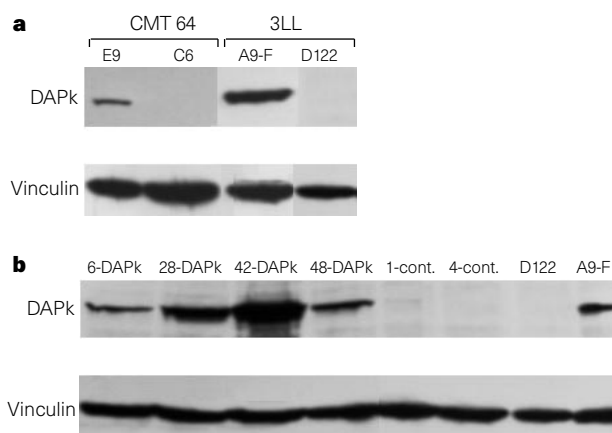


Figure 1 Immunoblot analysis of DAP kinase expression. **a**, DAP kinase protein was measured in low- (A9-F and E9) and high- (D122 and C6) metastatic clones derived from the 3LL and CMT64 tumours. **b**, Immunoblot analysis of parental D122 cells, and of the different G418-resistant derivative clones transfected with pcDNA control vector (-cont.) or with pcDNA-DAP kinase (+DAPk). The low-metastatic A9-F clone is used as a reference.

revealed that these cells had lost exogenous DAP kinase expression (Fig. 2d). Treatment of these recovered tumour cells with 5-aza-2'-deoxycytidine, an inhibitor of DNA methylation, completely restored DAP kinase protein expression (Fig. 2d), suggesting that attenuation of expression from the transgene was caused by DNA methylation.

A second *in vivo* selection was applied to the A9-F low-metastatic clone in an attempt to assess the status of the endogenous DAP kinase in the rare macroscopic lung lesions appearing late after its i.v. injection. The development of these metastatic lesions correlated with loss of DAP kinase expression (an example of three lung lesions appearing in a single mouse is shown in Fig. 2e). Altogether, the calculated value of loss of DAP kinase expression in these *in vivo* selections was 0.54 ($n = 11$). In contrast, random subcloning in culture of the A9-F cells showed that, without selection, the frequency of DAP kinase low expressors was 0.04 ($n = 48$) (the difference from the *in vivo* selections is significant at $P < 5 \times 10^{-4}$). Treatment of cells recovered from one of the lung lesions lacking DAP kinase, with 5-aza-2'-dioxycytidine restored protein expression to the normal levels of A9-F cells (Fig. 2e). DNA methylation was therefore also responsible for silencing the endogenous DAP kinase gene in the *in vivo*-selected A9-F cells, as previously documented for various tumour-suppressor genes¹⁷. Yet, this may not be an exclusive mechanism for suppressing DAP kinase expression

because demethylation failed to restore DAP kinase expression in the parental D122 high-metastatic clone. Moreover, in the CMT64 high-metastatic clone, gross DNA rearrangements of the DAP kinase gene could be identified by Southern blot analysis (not shown). We suggest that loss of DAP kinase expression (either from the transgene or from the endogenous gene) provides a positive selective advantage during the formation of lung metastases.

Next, the mechanisms underlying the suppressive effects of DAP kinase on metastasis and local tumour growth were studied. *In situ* TUNEL staining was performed on histological sections of footpad tumours. The apoptotic index in the slow-growing local tumours, formed by DAP-kinase-transfected cells, was higher than that in the tumour mass formed by the control clone (Fig. 3a). The mean values were $6.3\% \pm 1.13$ and $1.9\% \pm 0.35$ for 42-DAPk and the control clone, respectively ($P \ll 0.001$), reflecting a tremendous difference in total cell death in view of the rapid elimination of apoptotic cells in tissues¹⁸. These results implicated the DAP kinase gene in the augmentation of the threshold sensitivity of the tumour cells to apoptotic signals. We therefore exposed the transfected cells *in vitro* to two types of apoptotic stimuli, tumour-necrosis factor- α (TNF- α) and anchorage-independent cell growth. The Flag-tagged DAP kinase protein in clone 42-DAPk was tightly regulated by TNF- α , as reflected by its conversion into a form that rapidly migrates on gels (Fig. 3b, inset). At the phenotypic level, these transfectants displayed

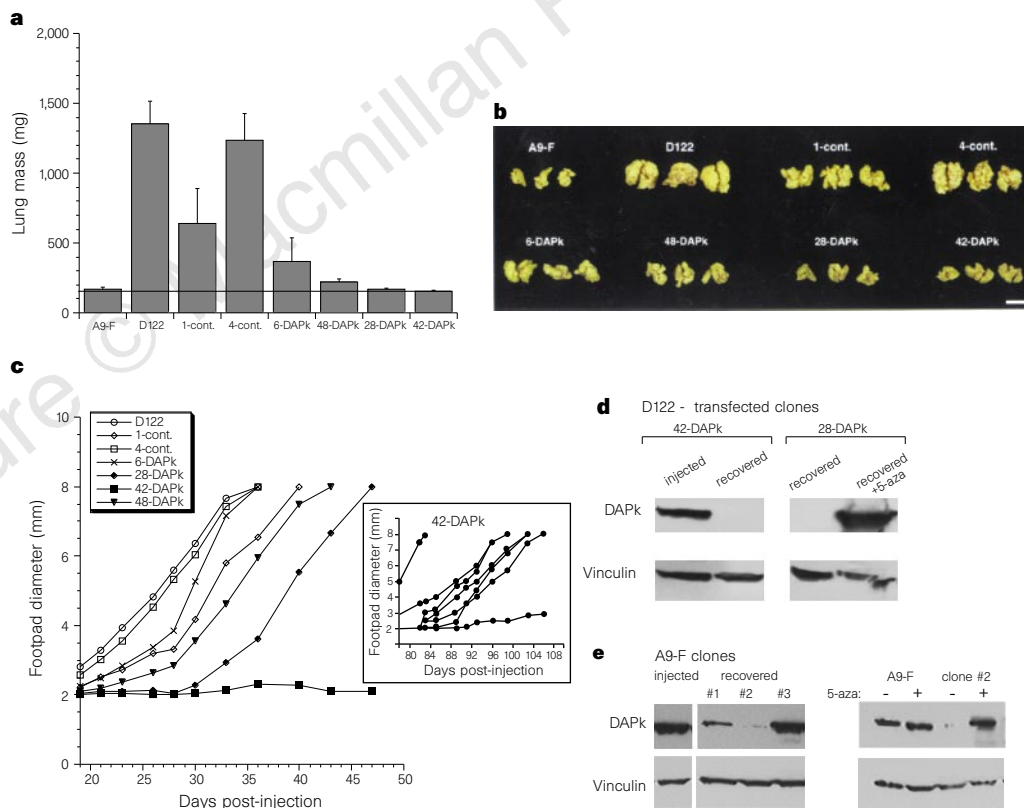


Figure 2 Inverse relationships between DAP kinase expression and metastatic activity. **a**, Transfection with DAP kinase strongly suppresses experimental metastasis. Values are mean lung weight $\pm$ s.d. of 5 individuals in each group. The solid line indicates the average lung weight of uninjected mice. Differences between the less aggressive 1-cont. clone and each one of the DAP kinase transfectants were significant at $P < 0.001$ for 48-, 28- and 42-DAPk clones, and at $0.025 < P < 0.05$ for the low-expressing clone 6-DAPk (the last clone differed from 4-cont. clone and parental D122 cells at $P < 0.001$). **b**, Three representative lungs from each group of mice as in **a**. Scale bar, 1 cm. **c**, Transfection with DAP-kinase delays the growth of local tumours. Values present the mean pad diameter of eight individuals in each group. The size difference between tumours formed by the slowest-growing control clone and by 28-DAPk clone was significant at

$P < 0.001$. The outburst of clone 42-DAPk after day 80 in each of 7 injected mice is shown in the inset. **d**, *In vivo* selections of DAP kinase revertants. Clone 42-DAPk was tested for protein expression both before the i.f.p. injections, and after the recovery in culture of spontaneous lung lesions appearing on day 35 post-surgery. Clone 28-DAPk was tested after recovery in culture from the spontaneous lung lesions without or with treatment with 5-aza-2'-deoxycytidine. **e**, *In vivo* selections of A9-F cells. DAP kinase expression was assessed in the original A9-F cells used for injection, and in cells recovered in culture from three independent lung lesions which appeared on day 61. Two out of three lesions display attenuated levels of DAP kinase, one of which is below detection limits. The latter clone was also treated with 5-aza-2'-deoxycytidine.

Table 1 Experimental metastasis of DAP kinase transfectants

Clone	D122	1-cont.	4-cont.	6-DAPk	48-DAPk	28-DAPk	42-DAPk
Number of mice that developed lung metastasis	5/5	5/5	5/5	5/5	5/5	4/5	0/5
Mean number of metastatic lesions per mouse*	>100	>100	>100	34 ± 27	12 ± 7	3 ± 2	0

* The number of metastatic lesions was determined by counting surface nodules under a binocular microscope.

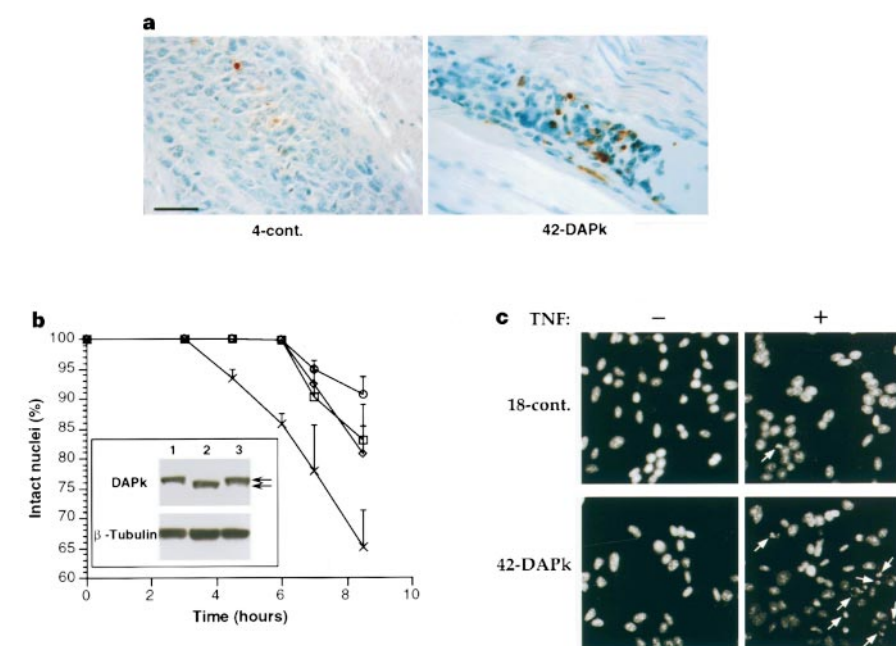


Figure 3 a, *In situ* TUNEL staining of footpad sections on five days after local injection of 4-cont. cells (left) or 42-DAPk cells (right). Scale bar, 100 μ m. **b**, Kinetics of the killing by TNF- α plus cycloheximide. The 42-DAPk transfectants (cross) were compared to the parental D122 cells (circle) and to 4-cont. (square) and 18-cont. (diamond). The values are the mean of the percentage of intact nuclei $\pm$ s.d. The inset shows immunoblot analysis of DAP kinase protein, from clone 42-DAPk, incubated for 4 h with no treatment, treatment with TNF- α and cycloheximide, or treatment with cycloheximide alone (lanes 1–3, respectively). **c**, DAPI staining of nuclei before and after treatment with TNF- α . The 18-cont. and 42-DAPk transfectants were treated with a combination of TNF- α and cycloheximide (right panels, marked by a plus), or with cycloheximide alone (left panels, marked by a minus). DAPI staining was done after 6 h. Arrows indicate apoptotic nuclei.

higher sensitivity to the apoptotic effects of TNF- α . Fragmented nuclei appeared much faster and total cell death was greater compared to the control D122 clones that lacked DAP kinase (Fig. 3b, c). A second type of apoptotic stress was induced by growing cells under anchorage-independent conditions in soft agar. In contrast to the parental D122 cells and control clones which formed large viable colonies in semi-solid medium, the various DAP-kinase-transfected clones formed small colonies of dying cells (Fig. 4A and B, a, b). Most of the cells in these colonies had an apoptotic morphology, and contained fragmented nuclei (compare Fig. 4B, c and d and their insets). DAP kinase expression, therefore, also mediates cell death that is induced by anchorage-independent growth^{19–22}. Interestingly, the revertants of clone 42-DAPk, which formed lung lesions and did not express the transgene, lost their increased sensitivity to TNF- α (Fig. 4C, left) as well as the cell-death responses to detachment from extracellular matrix (Fig. 4C, right). Based on these experiments, we suggest that DAP-kinase-mediated suppression of metastasis results, at least in part, from increased apoptotic sensitivity to various death-inducing stimuli.

We have excluded a direct involvement of cytotoxic T cells in the DAP-kinase-mediated suppression of metastasis. DAP-kinase-transfected clones did not display increased sensitivity to killing by cytotoxic T cells induced against immunogenic K^b D122 transfectants²³, and the cell-surface major histocompatibility complex (MHC) class I expression remained low, as previously reported for the parental D122 cells¹⁵ (not shown).

In conclusion, we have made two important findings. First, we have demonstrated the existence of inverse relationships between DAP kinase expression and the ability of carcinoma cells to metastasize to the lungs. Loss of DAP kinase expression therefore may convey selective advantage during the transition of lung carcinoma tumour cells from a low- to a more aggressive high-metastatic phenotype. Second, our results suggest a mechanism for this selective process, by coupling the loss of DAP kinase expression

to resistance to various apoptotic stimuli encountered by tumour cells, particularly during the multiple stages of metastasis. These findings strongly support the notion that loss of apoptotic control is an important factor in metastasis. Further support came from the finding that the dormancy of micrometastases is caused by indirect stimulation of apoptosis by angiogenic inhibitors²⁴, and from proof of a link between experimental metastasis and loss of p53 (ref. 25) or CC3 (ref. 26). We propose that DAP kinase may control apoptosis at early stages of metastasis, such as detachment from the original tumour (for example, cell death that is induced by loss of cell-matrix interactions, named anoikis) and transport in the circulation. In the bloodstream, metastasizing cells may be killed by interactions with monocytes, natural killer cells, and neutrophils, by exposure to cytokines, to nitric oxide anions, or by different mechanical stresses^{27–30}. This high number of apoptotic stimuli may explain why restoration of DAP kinase was much more effective in suppressing metastasis after the intravenous injections than in reducing local tumour growth in the foreign environment of footpad muscles. We suggest that loss of DAP kinase expression may act with the many other genetic alterations affecting cellular motility, immunity or invasion in the multistep process of metastasis. □

Methods

Growth conditions, transfections and immunoblot analysis. The 3LL clone-D122 and the CMT64 clone -C6 are the high-metastatic clones and clones A9-F, and E9 are the low-metastatic counterparts, respectively. All cells were grown in DMEM medium supplemented with 10% fetal calf serum (Gibco BRL). D122 cells were stably transfected with pcDNA3-DAP kinase and selected with G418 (800 μ g ml⁻¹) as described². Cell lysates, always prepared from exponentially growing cells, were analysed on immunoblots with anti-DAP-kinase monoclonal antibodies, anti-vinculin, and anti- β -tubulin antibodies (Sigma) as described². Clones 1-DAPk and 21-DAPk expressed exogenous DAP-kinase protein at levels similar to clone 28-DAPk (data not shown).

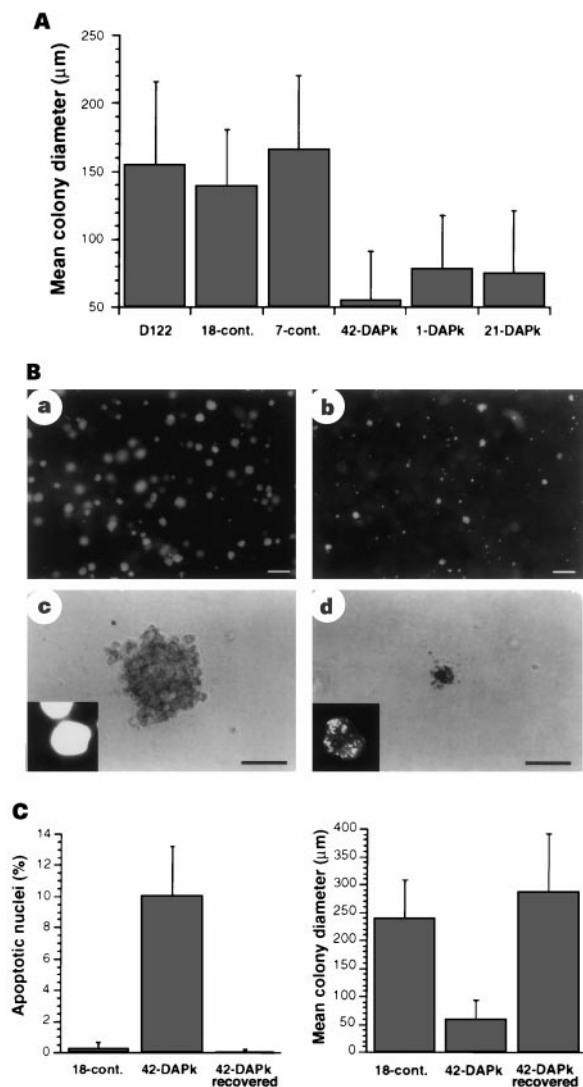


Figure 4 The growth of the D122 transfectants in semi-solid medium. **A**, Values are the mean colony diameter of 100 clones from each group $\pm$ s.d. The difference between the controls (such as 18-cont.) and the DAP kinase transfectants (such as 1-DAPk) was significant at $P \ll 0.001$. **B**, Light microscopy of the clones cultured in soft agar for 7 days, comparing the parental D122 cells (left: **a**, **c**) to H2-DAPk cells (right: **b**, **d**). Bars correspond to 350 μ m in the upper panels (**a**, **b**) and to 80 μ m in the lower panels (**c**, **d**). The insets in **c** and **d** show DAPI staining of nuclei from the corresponding cells. **C**, *In vivo* selection for attenuated DAP kinase expression ablates the increased sensitivity of clone 42-DAPk to TNF- α (left) and to death by detachment from extracellular matrix (right). Responses to TNF- α were assayed as for Fig. 3b at 6h. The original 42-DAPk clone was compared to the cultures recovered from the spontaneous lung metastases.

Tumour development and metastasis. All mice injected were 10–12-week-old C57BL/6 females. For experimental metastasis, the different D122-transfected clones were injected into tail veins (5×10^5 cells per mouse). Mice were killed 30–32 days later (when 50% of mice injected with parental D122 cells had died from lung metastasis), and their lungs were removed, weighed and fixed in Bouin's solution. Diameters of tumour-bearing feet, after intrafootpad injections (2×10^5 cells per mouse), were measured using calipers every 1–3 days. When tumour diameter reached 8–9 mm, tumour-bearing feet were amputated below the knee and the day of death resulting from spontaneous lung metastasis was scored for each individual mouse, in some cases, the tumour cells were recovered in culture from dissected lung lesions. For the *in vivo* selections, A9-F cells were injected (10^6 cells per mouse) into the tail veins of γ -irradiated mice (500 rad for 10 min), and killed after 2 months. Individual

lung lesions were removed, and the tumour cells were recovered in culture. Treatment of cell cultures with 5-aza-2'-deoxycytidine (10 μ M) (Sigma) was done by exposing the cells to the drug for 24 h during the exponential phase of growth. Protein expression was measured 72 h later.

In situ TUNEL assay. Fragments of mouse footpads were fixed for 12 h in 4% buffer formaldehyde (Frutarom), embedded in paraffin and sectioned (4 μ m thick). TUNEL assay on these sections was performed according to the manufacturer's instructions (ApopTag Plus Peroxidase kit; Oncor, Gaithersburg). Six different sections were scored; in each case 500–1,000 tumour cells were counted and the mean apoptotic index was calculated.

TNF- α assays. Exponentially growing cells were treated with a combination of murine TNF- α (100 ng ml $^{-1}$; R&D systems, Minneapolis) and cycloheximide (5 μ g ml $^{-1}$; Sigma), or with cycloheximide alone, and assessed for apoptotic nuclei by DAPI staining as in ref. 2. Five different fields, each consisting of 100 nuclei, were scored for every time point.

Soft agar assays. The different clones were cultured in a semi-solid medium containing 0.33% soft agar (Bacto-agar; Difco) at an initial count of 5×10^3 cells per 6-cm plate, on top of a layer containing 0.5% agar. The diameters of the clones that appeared on day 7 were measured under a light microscope. Cells were recovered from the upper agar layer by dissolving the agar with 6M sodium iodide (10 min at 37 $^{\circ}$ C). The washed cells were assessed for apoptotic nuclei by DAPI staining.

Statistics. Unpaired one-tailed Student's *t*-test was performed in all statistical analyses except for the *in vivo* selection experiments of the low-metastatic A9-F cells, in which both the χ^2 test for two independent samples and the Fisher exact probability test were used to calculate the *P*-value.

Received 14 July; accepted 20 August 1997.

- Deiss, L. P., Feinstein, E., Berissi, H., Cohen, O. & Kimchi, A. Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. *Genes Dev.* **9**, 15–30 (1995).
- Cohen, O., Feinstein, E. & Kimchi, A. DAP-kinase is a Ca $^{2+}$ /calmodulin-dependent, cytoskeletal-associated protein kinase, with cell death-inducing functions that depend on its catalytic activity. *EMBO J.* **16**, 998–1008 (1997).
- Feinstein, E., Wallach, D., Boldin, M., Varfolomeev, E. & Kimchi, A. The death domain: a module shared by proteins with diverse cellular functions. *Trends Biochem. Sci.* **20**, 342–344 (1995).
- Deiss, L. P. & Kimchi, A. A genetic tool used to identify thioredoxin as a mediator of a growth inhibitory signal. *Science* **252**, 117–120 (1991).
- Clarke, A. R. *et al.* Thymocyte apoptosis induced by p53-dependent and independent pathways. *Nature* **362**, 849–852 (1993).
- Lowe, S. W., Schmitt, E. M., Smith, S. W., Osborne, B. A. & Jacks, T. p53 is required for radiation-induced apoptosis in mouse thymocytes. *Nature* **362**, 847–849 (1993).
- Haffner, R. & Oren, M. Biochemical properties and biological effects of p53. *Curr. Opin. Genet. Dev.* **5**, 84–90 (1995).
- Graeber, T. G. *et al.* Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**, 88–91 (1996).
- Morin, P. J., Vengenstein, B. & Kinzler, K. W. Apoptosis and APC in colorectal tumorigenesis. *Proc. Natl Acad. Sci. USA* **93**, 7950–7954 (1996).
- Bakhshi, A. *et al.* Cloning the chromosomal breakpoint of t(14;18) human lymphomas: clustering around JH on chromosome 14 and near a transcriptional unit on 18. *Cell* **41**, 899–906 (1985).
- Tsujimoto, Y. *et al.* Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the t(11;14) chromosome translocation. *Nature* **315**, 340–343 (1985).
- Reed, J. C. Regulation of apoptosis by bcl-2 family proteins and its role in cancer and chemoresistance. *Curr. Opin. Oncol.* **7**, 541–546 (1995).
- Naik, P., Karmir, J. & Hanahan, D. The rise and fall of apoptosis during multistage tumorigenesis: down-modulation contributes to tumor progression from angiogenic progenitors. *Genes Dev.* **10**, 2105–2116 (1996).
- Kissil, J. L. *et al.* DAP-kinase loss of expression in B-cell leukemia and bladder carcinoma cell lines: possible implications for role as tumor suppressor gene. *Oncogene* **15**, 403–407 (1997).
- Eisenbach, L., Segal, S. & Feldman, M. MHC imbalance and metastatic spread in Lewis lung carcinoma clones. *Int. J. Cancer* **32**, 113–120 (1983).
- Layton, M. G. & Franks, L. M. Heterogeneity in a spontaneous mouse lung carcinoma: selection and characterization of stable metastatic variants. *Br. J. Cancer* **49**, 415–421 (1984).
- Laird, P. W. Oncogenic mechanisms mediated by DNA methylation. *Mol. Med. Today* **3**, 223–229 (1997).
- Bursch, W., Paffe, S., Putz, B., Barthel, G. & Schulte-Hermann, R. Determination of the length of the histological stages of apoptosis in normal liver and in altered hepatic foci of rats. *Carcinogenesis* **11**, 847–853 (1990).
- Meredith, J. E. Jr, Fazeli, B. & Schwartz, M. A. The extracellular matrix as a well survival factor. *Mol. Biol. Cell* **4**, 953–961 (1993).
- Meredith, J. E. Jr & Schwartz, M. A. Integrins, adhesion and apoptosis. *Trends Cell Biol.* **7**, 146–150 (1997).
- Frisch, S. M. & Francis, H. Disruption of epithelial cell-matrix interactions induces apoptosis. *J. Cell Biol.* **124**, 619–626 (1994).
- Boudreau, N., Sympon, C. J., Werb, Z. & Bissell, M. J. Suppression of ICE and apoptosis in mammary epithelial cells by extracellular matrix. *Science* **267**, 891–893 (1995).
- Plaksin, S., Gelber, C., Feldman, M. & Eisenbach, L. Reversal of the metastatic phenotype in Lewis lung carcinoma cells after transfection with syngeneic H-2Kb gene. *Proc. Natl Acad. Sci. USA* **85**, 4463–4467 (1988).
- Holmgren, L., O'Reilly, M. S. & Folkman, J. Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression. *Nature Med.* **1**, 149–153 (1995).
- Nikiforov, M. A. *et al.* p53 modulation of anchorage independent growth and experimental metastasis. *Oncogene* **13**, 1710–1720 (1996).

26. Shtivelman, E. A link between metastasis and resistance to apoptosis of variant small-cell lung carcinoma. *Oncogene* **14**, 2167–2173 (1997).
27. Fidler, I. J. & Balch, C. M. The biology of cancer metastasis and implications for therapy. *Curr. Prob. Surg.* **24**, 129–209 (1987).
28. Weiss, L. Biochemical interactions of cancer cells with the microvasculature during hematogenous metastasis. *Cancer Metastasis Rev.* **11**, 227–235 (1992).
29. Xie, K. P., Dong, Z. Y. & Fidler, I. J. Activation of nitric oxide synthase gene for inhibition of cancer metastasis. *J. Leuk. Biol.* **59**, 797–803 (1996).
30. Yamauchi, A., Taga, K., Mostowski, H. S. & Bloom, E. T. Target cell-induced apoptosis of interleukin-2-activated human natural killer cells: roles of cell surface molecules and intracellular events. *Blood* **87**, 5127–5135 (1996).

Acknowledgements. We thank E. Feinstein for discussions and for comments on the manuscript, J. L. Kissil, S. Ben-Bassat, D. Popovic and H. El-Shami for technical assistance, and L. M. Franks for the CMT-64 clones. The work was supported by QBI Ltd and by a grant from the Israeli Ministry of Science. A.K. is an Incumbent of the Helena Rubinstein Chair of Cancer Research. L.E. is an Incumbent of the G.F. Duckwitz Chair of Cancer Research.

Correspondence and requests for materials should be addressed to A.K. (e-mail: LVKIMCHI@WEIZMANN. weizmann.ac.il).

Herpes viral cyclin/Cdk6 complexes evade inhibition by CDK inhibitor proteins

Charles Swanton*, David J. Mann*, Bernhard Fleckenstein†, Frank Neipel†, Gordon Peters* & Nic Jones*

* Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

† Institut für Klinische und Molekulare Virologie, Friedrich-Alexander Universität, Loschgestrasse 7, D-8520 Erlangen, Germany

The passage of mammalian cells through the restriction point into the S phase of the cell cycle is regulated by the activities of Cdk4 and Cdk6 complexed with the D-type cyclins and by cyclin E/Cdk2 (refs 1–3). The activities of these holoenzymes are constrained by CDK inhibitory proteins^{4,5}. The importance of the restriction point is illustrated by its deregulation in many tumour cells^{6,7} and upon infection with DNA tumour viruses⁸. Here we describe the properties of cyclins encoded by two herpesviruses, herpesvirus saimiri (HVS) which can transform blood lymphocytes⁹ and induce malignancies of lymphoid origin in New World primates^{9,10}, and human herpesvirus 8 (HHV8) implicated as a causative agent of Kaposi's sarcoma and body cavity lymphomas^{11,12}. Both viral cyclins form active kinase complexes with Cdk6 that are resistant to inhibition by the CDK inhibitors p16^{Ink4a}, p21^{Cip1} and p27^{Kip1}. Furthermore, ectopic expression of a viral cyclin prevents G1 arrest imposed by each inhibitor and stimulates cell-cycle progression in quiescent fibroblasts. These results suggest a new mechanism for deregulation of the cell cycle and indicate that the viral cyclins may contribute to the oncogenic nature of these viruses.

Because transformation by DNA tumour viruses is accompanied by loss of restriction point control, we investigated the possibility that the HVS- and HHV8-encoded cyclins^{13,14} (V- and K-cyclin, respectively) could promote such deregulation and hence contribute to the known and suspected transforming properties of HVS and HHV8. We prepared cell lysates from Sf9 cells infected with recombinant baculoviruses expressing either cyclin together with Cdk6, a CDK subunit which is abundantly expressed in lymphocytes¹⁵ and shown to form active kinase complexes with the viral cyclins^{13,14}. The complexes were assayed for their ability to phosphorylate pRb in the absence or presence of bacterially expressed p21^{Cip1} or p27^{Kip1} inhibitor protein. Lysates containing Cdk6 and either of the viral cyclins efficiently phosphorylated pRb *in vitro* (Fig. 1a). Kinase activity was considerably higher than with lysates expressing cyclin D1 and Cdk6 (Fig. 1a), in agreement with previous results¹³. Importantly, both the V-cyclin- and K-cyclin-

associated activities were resistant to inhibition by either p21^{Cip1} or p27^{Kip1} over a 100-fold range of inhibitor concentration. This is in contrast to cyclin D1/Cdk6 activity which was sensitive to both inhibitors in a dose-dependent manner. A second family of CDK inhibitors typified by the p16^{Ink4a} protein also plays a significant role in restriction point control^{4,5}. This class of inhibitor is specific for Cdk4 or Cdk6 complexes. Both V-cyclin and K-cyclin complexes also showed resistance to a wide range of p16^{Ink4a} levels, whereas the cyclin D1/Cdk6 control was efficiently inhibited (Fig. 1b).

The difference in activity of the cyclin D1 and viral cyclin complexes and their sensitivity to inhibitors could not be explained by differences in the levels of cyclin/CDK binary complexes. ³⁵S-methionine labelling of infected cell extracts followed by immunoprecipitation with a Cdk6-specific antibody (Fig. 1c) showed comparable levels of immunoprecipitated cyclin D1 and K-cyclin and slightly higher levels of V-cyclin. We repeated the kinase inhibition assays using these Cdk6 immunoprecipitates. At the highest level of inhibitor, both the K-cyclin/Cdk6 and the V-cyclin/Cdk6 complexes were resistant to inhibition whereas the cyclin D1/Cdk6 complex was completely inhibited by p16^{Ink4a} and p27^{Kip1} and significantly inhibited by p21^{Cip1} (Fig. 1d). Therefore the results with the crude extracts and the immunoprecipitated complexes were directly comparable. To examine the sensitivity of the viral complexes further, we extended the ratio of inhibitor:complex by repeating the kinase assays with the highest level of inhibitor but 1/10 or 1/100 the level of immunoprecipitated Cdk6-containing complex. With a 10-fold decrease, partial inhibition of the K-cyclin complex was observed with p16^{Ink4a} and p27^{Kip1} but there was still little effect on the V-cyclin complex. With 100-fold less complex, although inhibition of the K- and V-cyclin activities was evident, significant activity remained (about 15% in the case of K-cyclin and 30% in the case of V-cyclin). Even at this dilution, however, p21^{Cip1} had negligible effect. Thus these results demonstrate that the viral cyclin complexes are significantly more resistant to inhibition than cyclin D1 complexes.

The efficiency of inhibition is likely to be defined by the binding affinity of the inhibitor for a given cyclin/CDK complex. The p21^{Cip1} and p27^{Kip1} inhibitors have been shown to bind both cyclin and CDK subunits^{16–19} and the interactions with both subunits contribute to the inhibition of the holoenzyme. We considered whether the resistance of the viral cyclin/Cdk6 complexes was due to lack of interaction between the viral cyclins and these inhibitors. Cyclins were synthesized by coupled transcription/translation in the presence of ³⁵S-methionine and mixed with unlabelled lysates containing p21^{Cip1}, p27^{Kip1} or Cdk6. The unlabelled components were immunoprecipitated and analysed for the presence of the labelled cyclin. Cyclin D1, K-cyclin and V-cyclin were all efficiently co-immunoprecipitated together with Cdk6 (Fig. 2a). Cyclin D1 was also immunoprecipitated with p21^{Cip1} and p27^{Kip1} inhibitors, confirming the direct binding shown previously using this assay (Fig. 2a)¹⁷. In marked contrast, p21^{Cip1} failed to co-immunoprecipitate either viral cyclin (Fig. 2a). Additionally, V-cyclin failed to interact with p27^{Kip1} (Fig. 2a) and although K-cyclin did show some interaction, it was 15-fold less than that of p27^{Kip1} with cyclin D1. The inefficient interaction was confirmed by co-expressing viral cyclin, Cdk6 and the p21^{Cip1} or p27^{Kip1} inhibitor in Sf9 cells in the presence of ³⁵S-methionine. The infected extracts were subjected to immunoprecipitation with an antibody specific to the inhibitor. Figure 2b shows efficient co-immunoprecipitation of cyclin D1 and Cdk6 with p21^{Cip1} or p27^{Kip1}. In contrast, at least 10-fold less viral cyclin was co-immunoprecipitated. These results demonstrate that the viral cyclin/Cdk6 complexes are resistant to inhibition by p21^{Cip1} and p27^{Kip1} because of the inability of these inhibitors to interact efficiently with the cyclin subunits. This conclusion is supported by the lack of conservation of certain key residues predicted to be crucial for the p27^{Kip1} interaction. The structure of cyclin A/Cdk2 bound to the N terminus of p27^{Kip1} has recently been described²⁰,

illustrating a region of p27^{Kip1} that binds directly to cyclin A through hydrophobic and electrostatic interactions. The surface residues of cyclin A lining the p27^{Kip1}-binding pocket are highly conserved among mammalian cyclins known to interact with p27^{Kip1}. The viral cyclins show less conservation in this region. Notably, E220 of cyclin A, which forms a hydrogen bond with S27 and a salt bridge with R30 of p27^{Kip1}, is not conserved (Fig. 2c). Thus the loss of electrostatic character in the predicted p27^{Kip1} binding surface of the viral cyclins may be responsible for their lack of inhibition by p27^{Kip1}. The mechanism behind the resistance of the viral cyclin complexes to p16^{Ink4a} is presently unknown and there is no evidence that this inhibitor interacts directly with the cyclin. We speculate that the resistance could arise from higher affinity of the viral cyclin for the Cdk6 subunit thus preventing its possible displacement by the inhibitor–Cdk6 interaction.

The results described above demonstrate that V-cyclin/Cdk6 and K-cyclin/Cdk6 complexes are resistant to inhibition by both families of inhibitor implicated in regulating the passage of cells from G1 into S. These properties suggest that they may stimulate the progression of cells through the G1 period of the cell cycle even in the presence of high levels of CDK inhibitor activity. We therefore assessed the ability of the K-cyclin to release a G1-arrest imposed by

the ectopic expression of p16^{Ink4a}, p21^{Cip1} or p27^{Kip1} (Table 1). The human osteosarcoma cell line U2-OS was transiently transfected with a plasmid encoding a CDK inhibitor together with a plasmid encoding the cell-surface marker CD8 to enable specific analysis of the transfected cells. Cells expressing each of the inhibitors accumulated in the G1 phase of the cell cycle. Co-transfection of K-cyclin partially, or in some cases fully, prevented the p16^{Ink4a}-imposed arrest whereas cyclin D1 had little effect. In p21^{Cip1}-arrested cells, K-cyclin expression not only prevented arrest but significantly decreased the percentage of cells present in the G1 phase compared to the untransfected controls. Overexpression of cyclin D1 largely prevented p21^{Cip1}-imposed arrest although there was no reduction in the G1 population compared to the control. The ability of cyclin D1 to overcome a p21^{Cip1}-mediated arrest has been previously reported¹⁹. Overexpression of K-cyclin also efficiently prevented a p27^{Kip1}-imposed G1 arrest whereas cyclin D1 was unable to do so. Western analysis of extracts prepared from the transfected population of cells indicated that both cyclin D1 and K-cyclin were efficiently expressed (data not shown).

Accumulating evidence suggests a separate role for both cyclin D1 and cyclin E complexes in promoting the transition of cells from the G1 phase into the S phase^{21–24}. It is noteworthy that the K-cyclin

Table 1 Ability of K-cyclin to release G1 arrest

Experiment	Inhibitor:	Increase in G1 (%)								
		p16			p21			p27		
	Cyclin:	–	Cyclin D1	K-cyclin	–	Cyclin D1	K-cyclin	–	cyclin D1	K-cyclin
1		41.2	40.6	11.6	30.0	5.7	–26.0	41.3	35.7	9.8
2		21.6	18.0	–6.5	19.1	6.4	–30.1	21.4	20.8	–1.8
3		45.4	39.0	13.4	13.7	–4.2	–30.3	20.1	17.6	–9.1
4		25.1	23.5	–5.9	21.1	5.8	–24.6	25.0	17.8	4.4
5		55.6	43.3	10.0	23.8	7.4	–17.2	36.4	30.1	2.8

U2-OS cells were transfected with plasmids encoding the cell surface marker CD8, inhibitor protein (p16^{Ink4a}, p21^{Cip1} or p27^{Kip1} as indicated) and cyclin (cyclin D1 or K-cyclin as indicated). Data are represented as the percentage change in the number of cells in G1 phase of the cell cycle relative to the G1 population of cells transfected with the plasmid expressing CD8 only. The results from five independent experiments are shown.

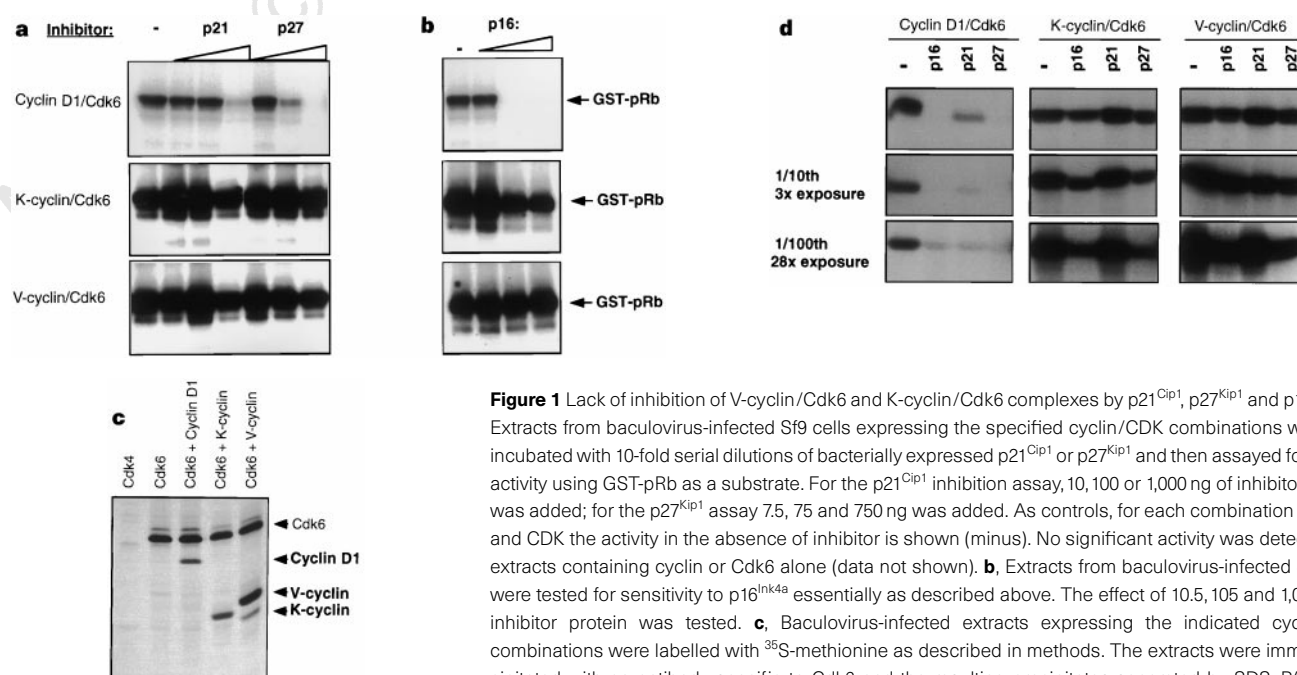


Figure 1 Lack of inhibition of V-cyclin/Cdk6 and K-cyclin/Cdk6 complexes by p21^{Cip1}, p27^{Kip1} and p16^{Ink4a}. **a**, Extracts from baculovirus-infected Sf9 cells expressing the specified cyclin/CDK combinations were pre-incubated with 10-fold serial dilutions of bacterially expressed p21^{Cip1} or p27^{Kip1} and then assayed for kinase activity using GST-pRb as a substrate. For the p21^{Cip1} inhibition assay, 10, 100 or 1,000 ng of inhibitor protein was added; for the p27^{Kip1} assay 7.5, 75 and 750 ng was added. As controls, for each combination of cyclin and CDK the activity in the absence of inhibitor is shown (minus). No significant activity was detectable in extracts containing cyclin or Cdk6 alone (data not shown). **b**, Extracts from baculovirus-infected Sf9 cells were tested for sensitivity to p16^{Ink4a} essentially as described above. The effect of 10.5, 105 and 1,050 ng of inhibitor protein was tested. **c**, Baculovirus-infected extracts expressing the indicated cyclin/CDK combinations were labelled with ³⁵S-methionine as described in methods. The extracts were immunoprecipitated with an antibody specific to Cdk6 and the resulting precipitates separated by SDS-PAGE and visualized by autoradiography. **d**, Complexes immunoprecipitated with an antibody against Cdk6 were tested for pRb phosphorylation activity in the presence of 1,000 ng p21^{Cip1}, 750 ng p27^{Kip1} or 1,050 ng p16^{Ink4a}. In lower panels, the amount of complex was reduced 10-fold or 100-fold, respectively, and assayed with the levels of inhibitor indicated above. The autoradiographs were exposed for different times to account for the different levels of kinase activity resulting from the dilutions.

complex can promote entry into S phase in the presence of high levels of each of the three inhibitors tested. This would imply that the viral cyclin/CDK complexes may be able to perform the role of both G1 cyclin complexes.

DNA tumour viruses have evolved mechanisms to force quiescent cells to enter S phase creating an environment more conducive to viral replication. To investigate whether the viral cyclin can perform such a role, an NIH 3T3 cell derivative expressing K-cyclin in an isopropyl- β -D-thiogalactoside (IPTG)-inducible manner was generated. In the absence of inducer, these cells efficiently quiesced upon serum deprivation as judged by BrdU labelling and fluorescence-activated cell sorting (FACS) analysis with roughly 90% of cells in the G0/G1 states (Fig. 3a). However, subsequent addition of IPTG to induce K-cyclin expression resulted in a significant propor-

tion of the cells entering S phase as demonstrated by the incorporation of BrdU into DNA. The increased expression of K-cyclin following induction was confirmed by western analysis (Fig. 3b). Thus, expression of K-cyclin is sufficient to induce S phase entry in quiescent cells.

In this report we highlight important biochemical and functional differences between the viral cyclins and the D-type cyclins, the most significant being their profound resistance to both families of CDK inhibitor proteins. The expression of such cyclins provides a new mechanism for deregulating cell-cycle progression. Such deregulation is an important prerequisite for efficient replication of many DNA tumour viruses and is essential for their tumorigenic potential. Given these parallels, we suggest that the viral cyclins are likely to be important for the replication and transforming properties of these herpesviruses. □

Methods

Plasmid and recombinant baculovirus constructs. Cyclin D1, HHV8 K-cyclin and HVS V-cyclin were cloned as polymerase chain reaction (PCR) fragments into the pRSETA vector (Invitrogen) as *Bam*H1 5'/*Eco*R1 3' fragments. In the case of HHV8 cyclin, the *Bam*H1 site at +36 was eliminated using a mutated 5' PCR oligonucleotide; the mutation had no effect on the amino-acid sequence of the cyclin. In some cases, single haemagglutinin (HA) epitopes were added to the C terminus by incorporating the epitope coding region in the 3' PCR primer. N-terminal 2HA-K-cyclin was made by ligating a 2HA coding oligonucleotide into the *Nco*I site at the 5' end of K-cyclin. All clones were sequenced.

p21^{Cip1}, p27^{Kip1} and p16^{Ink4a} clones in pRSETA and Cdk6 in Bluescript¹⁷ and the mammalian expression constructs containing CD8, p16^{Ink4a}, p21^{Cip1} and p27^{Kip1} (ref. 22) have been described previously. For mammalian expression constructs, coding sequences of the HA-epitope-tagged V-cyclin and K-cyclin sequences were subcloned from pRSETA vectors into the *Bam*H1/*Eco*R1 sites of

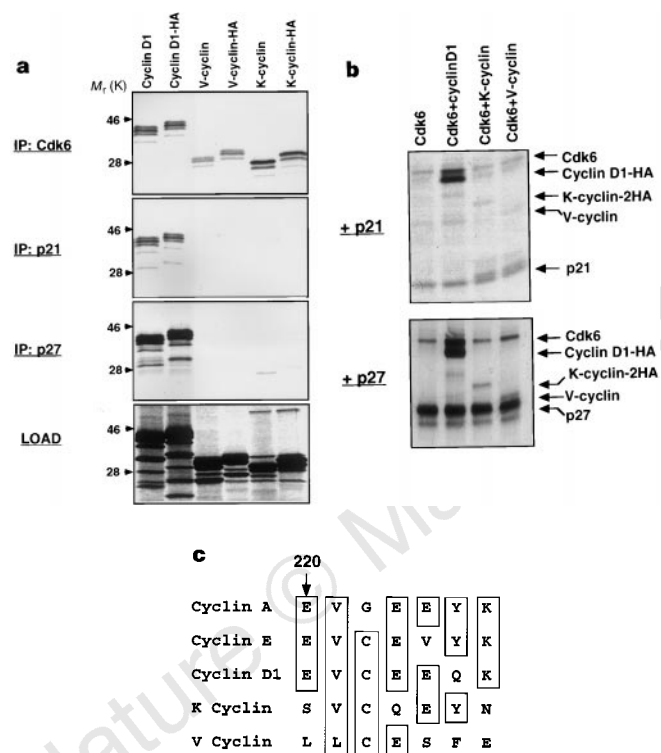


Figure 2 Co-precipitation of cyclins with p21^{Cip1}, p27^{Kip1} and Cdk6. ³⁵S-labelled cyclins were synthesized by *in vitro* translation and analysed by SDS-PAGE. For each cyclin, both an epitope-tagged product and an untagged product were analysed. **a**, Equal amounts of the labelled cyclins were mixed with translation lysates containing unlabelled Cdk6, p21^{Cip1} or p27^{Kip1} protein. Immunoprecipitations were performed using antibodies specific for the unlabelled component (as indicated) and the precipitates were analysed by SDS-PAGE and autoradiography to determine the proportion of labelled cyclin which was co-immunoprecipitated. **b**, Baculovirus-infected extracts expressing the indicated cyclin/Cdk combinations together with either p21^{Cip1} or p27^{Kip1} were labelled with ³⁵S-methionine as described in Methods. The extracts were immunoprecipitated with an antibody specific to p27^{Kip1} or p21^{Cip1} and the resulting precipitates separated by SDS-PAGE and visualized by autoradiography. In the case of K-cyclin, a double HA-tagged version was expressed because the mobility of K-cyclin is similar to that of p27. The levels of immunoprecipitated proteins were estimated using Molecular Dynamics phosphorimager analysis taking into account differing methionine content of the cyclin components. **c**, Sequence comparisons of K- and V-cyclins with the cyclins A, E and D1 that interact with p21^{Cip1} and p27^{Kip1}. The E220 residue of cyclin A which contributes to the electronegative surface of the cyclin and is involved in hydrogen bonding with p27^{Kip1} (ref. 20) is aligned with residues S60 and L61 of K- and V-cyclin, respectively.

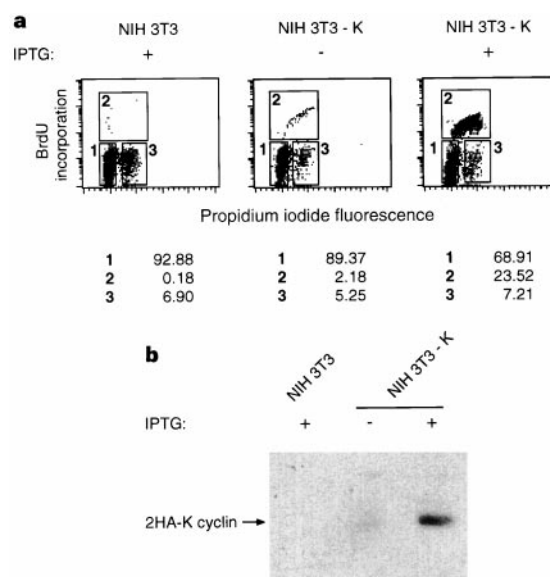


Figure 3 Stimulation of quiescent cells by expression of K-cyclin. An NIH 3T3 cell derivative was generated that expresses K-cyclin upon the addition of the inducer IPTG (NIH3T3-K). **a**, NIH 3T3 or NIH3T3-K cells were quiesced in the absence of IPTG by incubation in medium containing 0.2% serum for 72 h. As appropriate, K-cyclin expression was then induced by addition of 5 mM IPTG. BrdU (10 μ M) was added 10 h later and the cells incubated for a further 24 h. The cells were then collected, fixed and stained both with propidium iodide and anti-BrdU-isothiocyanate conjugate and analysed by FACS. The percentage of cells in either G1 (box 1), S (box 2) or G2/M (box 3) are indicated. **b**, Cells treated as above were lysed in SDS-PAGE sample buffer and subjected to western blot analysis using an anti-HA antibody.

the mammalian expression vector pcDNA3 (Invitrogen). For recombinant baculovirus construction, the same fragments were subcloned into pvl1393 (Invitrogen). For the construction of the IPTG-inducible K-cyclin, HA-tagged K-cyclin was amplified from the pRSETA vector described above using oligonucleotide primers containing *NotI* restriction sites. The ensuing fragment was cloned into the *NotI* site of pOPRSVICAT (Stratagene) to give the plasmid DM179.

Recombinant *Autographa californica* nuclear polyhedrosis virus were produced using the BaculoGold transfection system (PharMingen) and subsequently amplified. Recombinant baculoviruses expressing cyclin D1 and cyclin E were a gift of C. Sherr, and Cdk6 a gift of M. Meyerson.

Antibodies and immunoblotting. Antisera generated against C-terminal peptides of human p21^{Cip1}, p27^{Kip1} and Cdk6 have been described in previous reports¹⁷. A polyclonal antibody against the C terminus of p27^{Kip1} (SC528; Santa Cruz) was used for ³⁵S immunoprecipitations from Sf9 cells. The 12CA5 anti-haemagglutinin antibody was from Boehringer Mannheim. Immunoblot analysis was performed as described²².

In vitro kinase assays. Sf9 cells were co-infected with the appropriate recombinant baculoviruses and whole-cell extracts prepared 48 h after infection as described²⁴. Phosphorylation reactions (40 µl) were performed in 50 mM Tris-HCl, pH 7.4, 10 mM MgCl₂, 1 mM dithiothreitol, 0.1 mM EGTA and incubated with 500 ng GST-pRb bound to glutathione-Sepharose beads, 100 µM ATP and 5 µCi [γ -³²P]ATP. After 5 min at 30 °C, reactions were terminated by the addition of 500 µl cold NETN (20 mM Tris-HCl, pH 8, 100 mM NaCl, 1 mM EDTA, 0.5% NP40). The beads were recovered, boiled in dissociation buffer and analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). For inhibition assays, the extracts were pre-incubated with an equal volume of histidine-tagged p16^{Ink4a}, p21^{Cip1} and p27^{Kip1} (10-fold dilutions) for 10 min at 30 °C.

For the immunoprecipitation kinase assays, extracts were prepared and immunoprecipitated with an antibody directed against Cdk6 essentially as described below. The resulting lysates were washed and resuspended in kinase buffer and kinase assays performed as described in ref. 15.

In vitro interaction assay. pRSET vectors containing the coding sequences of V-cyclin, K-cyclin and cyclin D1 were used to generate ³⁵S-labelled proteins by coupled transcription and translation using the TNT expression system (Promega). The same system was used to generate unlabelled Cdk6, p21^{Cip1} or p27^{Kip1} protein. Samples of each reaction were mixed, incubated for 30 min at 30 °C, diluted using cold 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1% NP40, 3% BSA and precipitated using antiserum against the unlabelled component. The labelled proteins in the immunoprecipitate were analysed by SDS-PAGE.

Cell culture, transfection and FACS analysis. Culture of U2-OS cells, transfection and analysis of the cell cycle distribution of transfected cells have been described²².

Co-infections of Sf9 cells have been described²⁵. For labelling of expressed proteins, culture medium was replaced 48 h after infection with 3 ml Grace's medium lacking methionine (Gibco) and containing 10% dialysed fetal calf serum. After 1 h the medium was replaced with a further 5 ml of medium lacking methionine and containing 10% dialysed serum with 140 µCi of ³⁵S-methionine (Amersham). Cells were left for 12 h and resuspended in 1 ml IP buffer (0.1% Tween-20, 50 mM Tris pH 7.5, 50 mM NaCl, 2 mM EGTA, 1 mM EDTA, 1 mM DTT, and protease inhibitors). The Cdk6 antibody LB02 (10 µl) was added to 500 µl of lysate in a total volume of 1 ml. After 1 h at 4 °C protein A-Sepharose beads were added, and the lysates rotated for a further 60 min. Beads were washed 3 times in IP buffer and resuspended in SDS-PAGE loading buffer before analysis on 12% gels. In the case of p21^{Cip1}/p27^{Kip1} immunoprecipitations, Sf9 cells were co-infected with viruses expressing inhibitor, cyclin and Cdk either individually or in combination, and labelling was performed as described above. RIPA buffer was used for the immunoprecipitation and subsequent washing.

To isolate IPTG-inducible cell lines, NIH3T3 cells were transfected with DM179 and p3'SS (Stratagene) expressing the *lacI* gene. G418- and hygromycin-resistant colonies were isolated and screened for IPTG-dependent K-cyclin expression by western analysis. For FACS analysis, BrdU-labelled cells were trypsinized and fixed in 70% ethanol for 30 min on ice. Detection of BrdU-incorporation with an anti-BrdU-isothiocyanate conjugate (1202 693, Boehringer Mannheim) was carried out according to the manufacturer's

protocol. Cells were stained for DNA with propidium iodide (10 µg ml⁻¹) and analysed using bivariate flow cytometry.

Received 14 May; accepted 26 August 1997.

- Sherr, C. J. Mammalian G1 cyclins. *Cell* **73**, 1059–1065 (1993).
- Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- Sherr, C. J. Cancer cell cycles. *Science* **274**, 1672–1677 (1996).
- Sherr, C. J. & Roberts, J. M. Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1163 (1995).
- Harper, J. W. & Elledge, S. J. Cdk inhibitors in development and cancer. *Curr. Opin. Genet. Dev.* **6**, 56–64 (1996).
- Hall, M. & Peters, G. Genetic alterations of cyclins, cyclin-dependent kinases, and cdk inhibitors in human cancer. *Adv. Cancer Res.* **68**, 67–108 (1996).
- Strauss, M., Lukas, J. & Bartek, J. Unrestricted cell cycling and cancer. *Nature Med.* **1**, 1245–1246 (1996).
- Nevins, J. R. Cell cycle targets of the DNA tumor viruses. *Curr. Opin. Genet. Dev.* **4**, 130–134 (1994).
- Biesinger, B. et al. Stable growth transformation of human T lymphocytes by herpesvirus saimiri. *Proc. Natl Acad. Sci. USA* **89**, 3116–3119 (1992).
- Fleckenstein, B. & Desrosiers, R. C. in *The herpesviruses* Vol. 1 (ed. Roizman, B.) 252–332 (Plenum, New York, 1982).
- Chang, Y. et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**, 1865–1869 (1994).
- Cesarmen, E. et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. *N. Engl. J. Med.* **332**, 1186–1191 (1995).
- Jung, J. U., Stager, M. & Desrosiers, R. C. Virus encoded cyclin. *Mol. Cell. Biol.* **14**, 7235–7244 (1994).
- Chang, Y. et al. Cyclin encoded by K5 herpesvirus. *Nature* **382**, 410 (1996).
- Meyerson, M. & Harlow, E. Identification of G1 kinase activity for cdk6, a novel cyclin D partner. *Mol. Cell. Biol.* **14**, 2077–2086 (1994).
- Toyoshima, H. & Hunter, T. p27, a novel inhibitor of G1 cyclin-Cdk protein kinase activity, is related to p21. *Cell* **78**, 67–74 (1994).
- Hall, M., Bates, S. & Peters, G. Evidence for different modes of action of cyclin-dependent kinase inhibitors: p15 and p16 bind to kinases, p21 and p27 bind to cyclins. *Oncogene* **11**, 1581–1588 (1995).
- Chen, J. et al. Cyclin binding motifs are essential for the function of p21^{Cip1}. *Mol. Cell. Biol.* **16**, 4673–4682 (1996).
- Lin, J., Reichner, C., Wu, X. & Levine, A. J. Analysis of wild-type and mutant p21 WAF-1 gene activities. *Mol. Cell. Biol.* **16**, 1786–1793 (1996).
- Russo, A. A., Jeffrey, P. D., Patten, A. K., Massague, J. & Pavletich, N. P. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* **382**, 325–331 (1996).
- Resnitzky, D. & Reed, S. I. Different roles for cyclin D1 and E in regulation of the G1-to-S transition. *Mol. Cell. Biol.* **15**, 3463–3469 (1995).
- Mann, D. J. & Jones, N. E2F-1 but not E2F-4 can overcome p16-induced cell cycle arrest. *Curr. Biol.* **6**, 474–483 (1996).
- Ohtsubo, M. et al. Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. *Mol. Cell. Biol.* **15**, 2612–2624 (1995).
- Roussel, R. et al. Rescue of defective mitogenic signalling by D-type cyclins. *Proc. Natl Acad. Sci. USA* **92**, 6837–6841 (1995).
- Kato, J. M. et al. Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4. *Genes Dev.* **7**, 331–342 (1993).

Acknowledgements. We thank C. Sherr and M. Meyerson for gifts of baculovirus recombinants. We thank M. Hall for technical advice on the cyclin-inhibitor binding assay and for providing purified inhibitor proteins. We also thank D. Davies and C. Hughes for help with FACS analysis, G. Clark and A. Davies for sequencing analysis, and G. Card and N. McDonald for advice and encouragement throughout the study. This study was funded by the Imperial Cancer Research Fund. C.S. is supported by University College London Medical School and E.N. by the 'Deutsche Krebsforschungsförderung' founded by Mildred Scheel foundation.

Correspondence and requests for materials should be addressed to N.J. (e-mail jonesn@icrf.icnet.uk).

Osmotic stress activates phosphatidylinositol-3,5-bisphosphate synthesis

Stephen K. Dove^{*†‡}, Frank T. Cooke^{‡§}, Michael R. Douglas^{||}, Lee G. Sayers[§], Peter J. Parker[§] & Robert H. Michell^{*†}

^{*} Centre for Clinical Research in Immunology and Signalling, and Departments of [†] Biochemistry and ^{||} Rheumatology, University of Birmingham, Birmingham B15 2TT, UK

[§] Imperial Cancer Research Fund Laboratories, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

[‡] These authors contributed equally to this work

Inositol phospholipids play multiple roles in cell signalling systems. Two widespread eukaryotic phosphoinositide-based signal transduction mechanisms, phosphoinositidase C-catalysed phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂) hydrolysis and 3-OH kinase-catalysed PtdIns(4,5)P₂ phosphorylation, make the second messengers inositol 1,4,5-trisphosphate (Ins(1,4,5)P₃) *sn*-1,2-diacylglycerol and PtdIns(3,4,5)P₃ (refs 1–7). In addition,

PtdIns(4,5) P_2 and PtdIns3P have been implicated in exocytosis and membrane trafficking⁸. We now show that when the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* are hyperosmotically stressed, they rapidly synthesize phosphatidylinositol-3,5-bisphosphate (PtdIns(3,5) P_2) by a process that involves activation of a PtdIns3P 5-OH kinase. This PtdIns(3,5) P_2 accumulation only occurs in yeasts that have an active *vps34*-encoded PtdIns 3-OH kinase, showing that this latter kinase makes the PtdIns3P needed for PtdIns(3,5) P_2 synthesis and

indicating that PtdIns(3,5) P_2 may have a role in sorting vesicular proteins. PtdIns(3,5) P_2 is also present in mammalian and plant cells: in monkey Cos-7 cells, its labelling is inversely related to the external osmotic pressure. The stimulation of a PtdIns3P 5-OH kinase-catalysed synthesis of PtdIns(3,5) P_2 , a molecule that might be a new type of phosphoinositide 'second messenger', thus appears to be central to a widespread and previously uncharacterized regulatory pathway.

While characterizing the phosphoinositides of [³H]inositol-

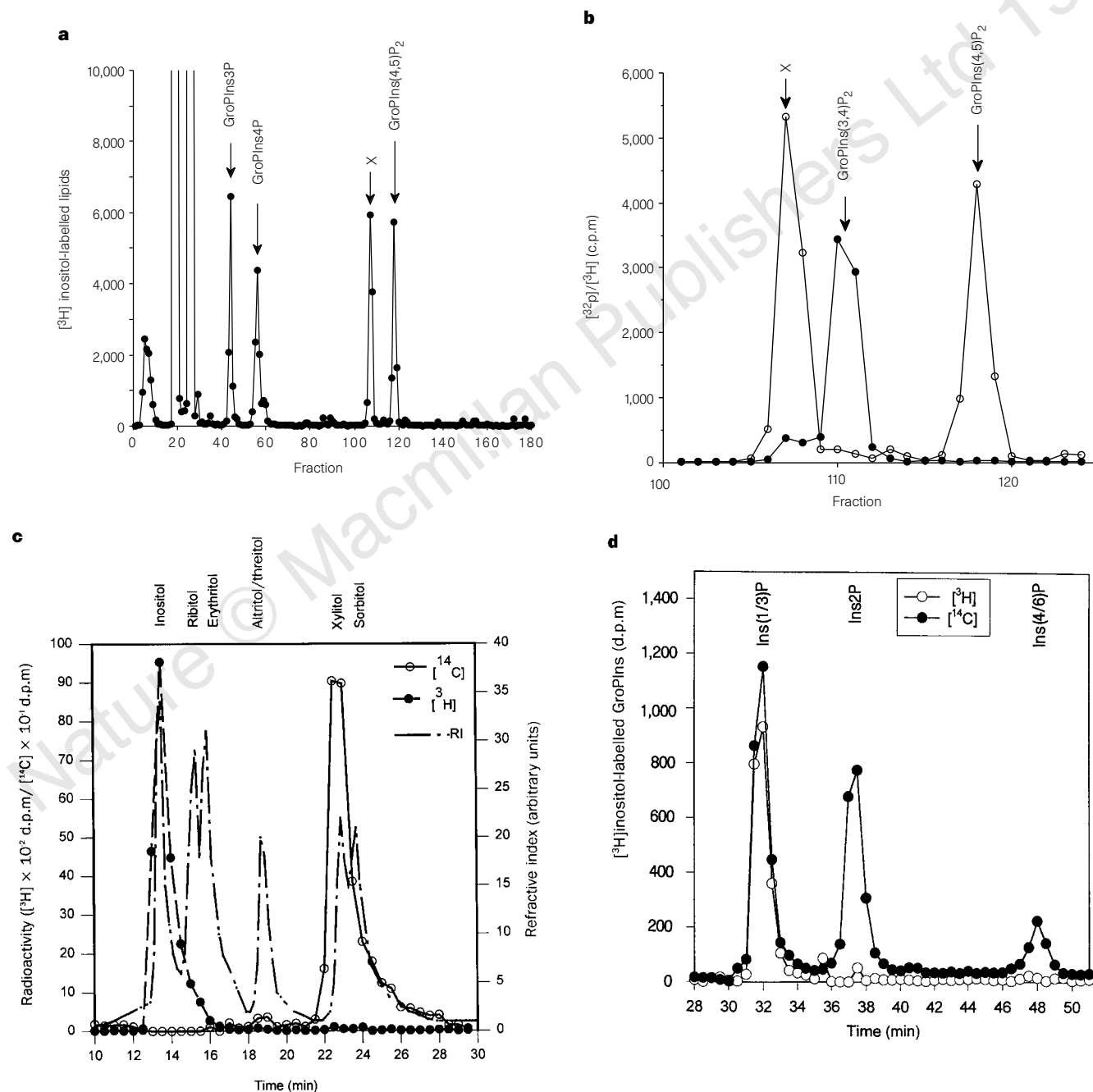


Figure 1 Identification of a novel yeast PtdIns(4,5) P_2 as PtdIns(3,5) P_2 . **a**, HPLC analysis of the deacylated [³H]inositol-labelled lipids of Zymolase-treated *S. cerevisiae* YPH499a shows the presence of similar amounts of GroPIns(4,5) P_2 and of a second putative GroPIns P_2 (labelled 'X'). **b**, The novel putative [³H]GroPIns P_2 ('X') and [³H]GroPIns(4,5) P_2 (circles) were both resolved from [³⁻³²P]GroPIns(3,4) P_2 (filled circles; [³H]PtdIns(3,4) P_2 was synthesized from labelled [³H]PtdIns4P with recombinant p85/p110 PI(3K)²⁷). **c**, HPLC separation of the polyols derived from simultaneous periodate oxidation of the [³H]inositol-

labelled Ins P_3 headgroup isolated from 'X' and of a ¹⁴C-labelled plant-derived Ins P_3 [Ins(3,4,5) P_2 and/or Ins(1,5,6) P_3]. RI denotes a refractive index trace of the simultaneous elution of unlabelled reference polyols. Two experiments gave similar results. **d**, Identification of the [³H]inositol-labelled GroPIns derived from 'X' as *sn*-3-glycerophospho-D-1-inositol. The InsP obtained by dephosphorylating and deglycerating [³H]GroPIns(3,5) P_2 was co-chromatographed with a [¹⁴C]-labelled Ins1P/Ins2P/Ins3P/Ins4P mixture (see Methods). The elution positions of reference inositol monophosphates are indicated.

labelled *S. cerevisiae* whose wall had been removed by Zymolase (ICN), we detected the deacylation product of a new [^3H]inositol-containing lipid ('X' in Fig. 1a, b). There was less 'X' in unperturbed cells (see below). 'X' had the charge of a deacylated PtdInsP₂ and was present at a concentration similar to deacylated PtdIns(4,5)P₂ (GroPIns(4,5)P₂). However, it eluted before either GroPIns(4,5)P₂ (Fig. 1a) or GroPIns(3,4)P₂ (deacylated [^{32}P]PtdIns(3,4)P₂; Fig. 1b). Following thin-layer chromatography (TLC) of lipids from [^3H]inositol-labelled and Zymolase-treated yeast, there were two labelled spots in the PtdInsP₂ region: deacylation and high-pressure liquid chromatography (HPLC) identified the faster-moving lipid as PtdIns(4,5)P₂, and the slower lipid yielded 'X'.

'X' is therefore derived from a PtdInsP₂ with a structure different from previously known PtdInsP₂ isomers. PtdIns(3,4)P₂ and PtdIns(3,4,5)P₃ are absent from yeast lipid extracts, so the PtdInsP₂ that yields X is unrelated to the PtdIns(4,5)P₂ 3-kinase signalling pathway. Formal identification of inositol trisphosphate (InsP₃) isomers can be achieved by characterizing the products of periodate oxidation^{9,10}, which opens the inositol ring between carbons bearing non-substituted vicinal hydroxyl groups. We therefore isolated the [^3H]inositol-labelled InsP₃ headgroup of the novel [^3H]PtdInsP₂ and mixed it with a natural [^{14}C]inositol-labelled *Spirodela polyrhiza* InsP₃ that yields xylitol on oxidation (Ins(3,4,5)P₃ and/or Ins(1,5,6)P₃ (C. Brearley, personal communication)). Following periodate treatment, reduction and dephosphorylation, the *Spirodela* [^{14}C]InsP₃ yielded [^{14}C]xylitol, confirming effective per-

iodate oxidation, but the [^3H]InsP₃ from the novel yeast lipid yielded unchanged [^3H] inositol (Fig. 1c). Of twenty InsP₃ isomers, only Ins(2,4,6)P₃ and Ins(1,3,5)P₃ are periodate-resistant. The novel GroPInsP₂ was converted to a GroPIns with alkaline phosphatase, and thence to an inositol monophosphate. This was Ins1P or its enantiomer Ins3P: no Ins2P or Ins4P/Ins6P was found (Fig. 1d). A lipid containing Ins(2,4,6)P₃ was thus eliminated, identifying the novel yeast lipid as PtdIns(3,5)P₂. There have earlier been hints pointing to the possible biological occurrence of PtdIns(3,5)P₂ (refs 11, 12), and PtdIns(3,5)P₂ has been independently reported in a fibroblast-like cell line¹³.

The fact that *S. cerevisiae* with cell walls weakened by Zymolase contain more PtdIns(3,5)P₂ than untreated cells suggested that stresses might enhance PtdIns(3,5)P₂ synthesis. We then observed that substantial hyperosmotic stresses (1.4 M sorbitol or 0.9 M NaCl) stimulate [^3H]PtdIns(3,5)P₂ accumulation in yeast in mid-log phase (Fig. 2a–c), but not in stationary-phase cells (not shown). The PtdIns(3,5)P₂ concentration rapidly rose seven- to twenty-fold after 5–15 min, reaching about the same concentration as PtdIns(4,5)P₂. It remained elevated for varying periods before returning towards the starting concentration, usually between 30 and 60 min (Fig. 2a, c). Lesser hyperosmotic stresses either provoked no PtdIns(3,5)P₂ accumulation (Fig. 2b) or caused smaller and more transient responses (not shown). While [^3H]PtdIns(3,5)P₂ was accumulating in the acutely stressed cells, the [^3H]PtdIns3P concentration usually decreased, with the net loss

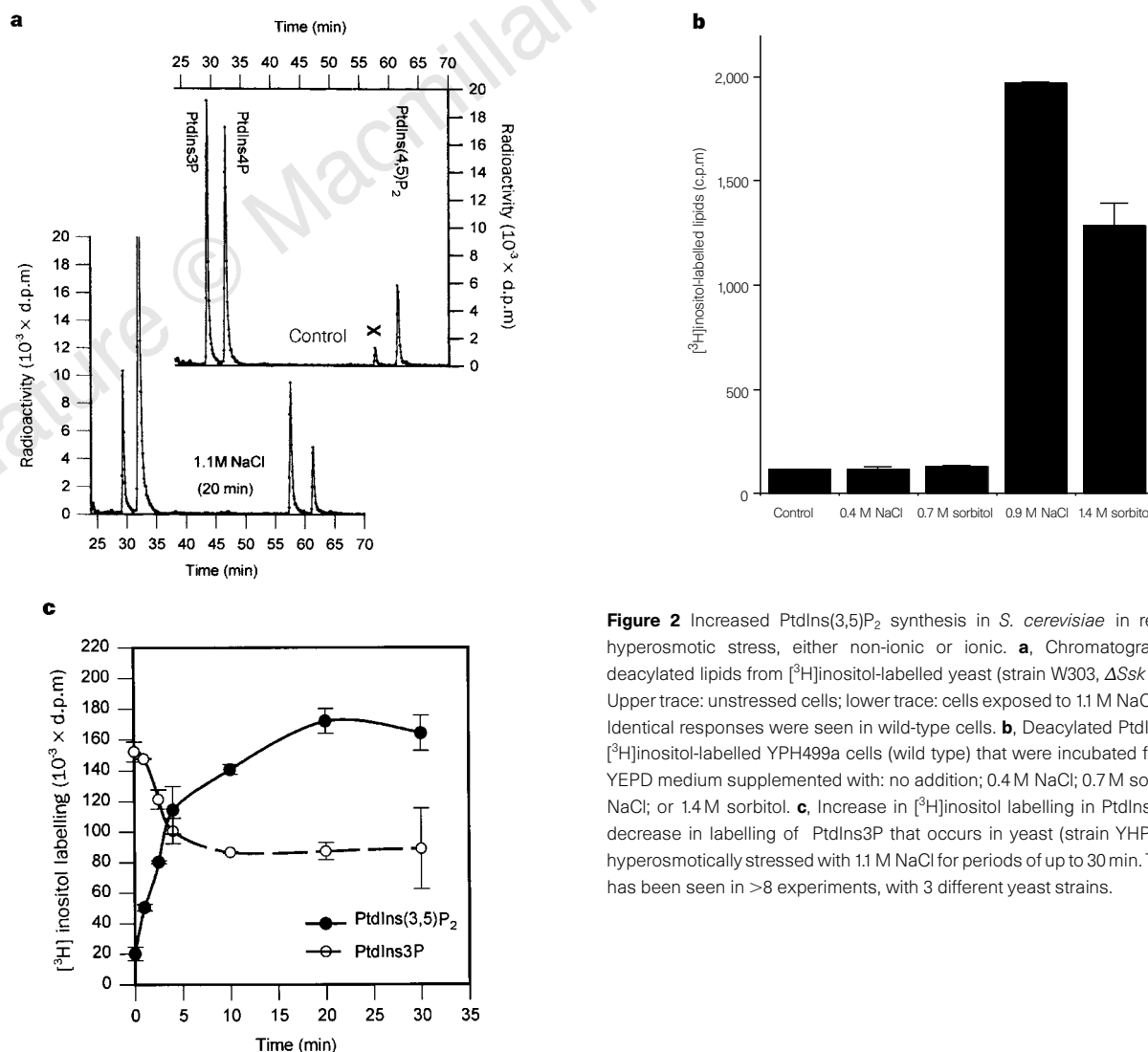


Figure 2 Increased PtdIns(3,5)P₂ synthesis in *S. cerevisiae* in response to hyperosmotic stress, either non-ionic or ionic. **a**, Chromatograms of the deacylated lipids from [^3H]inositol-labelled yeast (strain W303, ΔSsk1 ; see text). Upper trace: unstressed cells; lower trace: cells exposed to 1.1 M NaCl for 20 min. Identical responses were seen in wild-type cells. **b**, Deacylated PtdIns(3,5)P₂ of [^3H]inositol-labelled YPH499a cells (wild type) that were incubated for 10 min in YEPD medium supplemented with: no addition; 0.4 M NaCl; 0.7 M sorbitol; 0.9 M NaCl; or 1.4 M sorbitol. **c**, Increase in [^3H]inositol labelling in PtdIns(3,5)P₂ and decrease in labelling of PtdIns3P that occurs in yeast (strain YHP: wild type) hyperosmotically stressed with 1.1 M NaCl for periods of up to 30 min. This pattern has been seen in >8 experiments, with 3 different yeast strains.

of ^3H from PtdIns3P during the first few minutes sometimes accounting for up to half of the ^3H recruited into PtdIns(3,5)P₂ (Fig. 2a, c).

The fact that hyperosmotic stress triggers rapid PtdIns(3,5)P₂ synthesis suggests that PtdIns(3,5)P₂ might be an intracellular signalling molecule involved in controlling responses to some stresses. The best characterized responses of *S. cerevisiae* to hyperosmotic stresses are controlled through a pathway involving the MAP kinase homologue Hog1 (refs 14, 15). When we examined knockout mutants in three components of this signalling cascade (the SH3-domain-containing *Sho1* osmosensor, the *Ssk1* response regulator element of the *Shl1* two-component osmosensor, and the *Hog1* protein kinase¹⁵), they showed normal PtdIns(3,5)P₂ accumulation responses, indicating that the *Hog1* pathway does not

control the hyperosmotically evoked synthesis of PtdIns(3,5)P₂. Indeed, a ΔSsk1 mutant that shows a response identical to wild-type cells provided the data that illustrate a 'typical' PtdIns(3,5)P₂ response in Fig. 2a.

To define the synthetic route to PtdIns(3,5)P₂, we briefly labelled *S. cerevisiae* with [^{32}P]orthophosphate during rapid hyperosmotically stimulated PtdIns(3,5)P₂ accumulation. Under such labelling conditions, the last phosphate added to the inositol ring has the highest ^{32}P specific radioactivity¹⁰. To determine the distribution of ^{32}P between the 3- and 5-phosphate groups, [^{32}P]GroPIns(3,5)P₂ made from this [^{32}P]PtdIns(3,5)P₂ was mixed with [^3H]inositol-labelled GroPIns(3,5)P₂ from a separate batch of yeast and incubated with erythrocyte membranes in the absence of Mg²⁺ (when inositol polyphosphate 3-phosphatase is active but Ins(1,4,5)P₃ 5-phosphatase is inactive^{10,16}). During HPLC, the principal resulting [^{32}P]GroPInsP was well separated from GroPIns3P: it eluted in a position characteristic of GroPIns4P or GroPIns5P (not shown), identifying it as GroPIns5P. A further sample of the same [$^3\text{H}/^{32}\text{P}$]GroPIns(3,5)P₂ was incubated with alkaline phosphatase to remove both monoester phosphate groups.

The $^{32}\text{P}:^3\text{H}$ ratio of the doubly labelled GroPIns5P was ~80% of that of the initial lipid-derived GroPIns(3,5)P₂ sample, and the $^{32}\text{P}:^3\text{H}$ ratio of the GroPIns liberated by alkaline phosphatase was <0.1% of that of the PtdIns(3,5)P₂ (Table 1). Thus the phosphodiester phosphate at the 1-position was essentially unlabelled, and the 5-phosphate contained about four times as much ^{32}P as the 3-phosphate. This showed that the 5-phosphate is added last, and that PtdIns(3,5)P₂ is synthesized by 5-phosphorylation of PtdIns3P (which also shows rapid metabolic turnover). The transient depletion of PtdIns3P seen during stimulated PtdIns(3,5)P₂ synthesis (Fig. 2c) is as expected if continued PtdIns3P synthesis only partly compensates for rapid PtdIns3P consumption during PtdIns(3,5)P₂ synthesis (for a comparable situation, see ref. 1). This lends further credence to the pathway: PtdIns → PtdIns3P → PtdIns(3,5)P₂.

Recognition that PtdIns(3,5)P₂ is made from PtdIns3P by a 5-kinase suggested that the type III phosphoinositide 3-kinase encoded by *vps34* (refs 8, 17, 18) might supply the PtdIns3P for

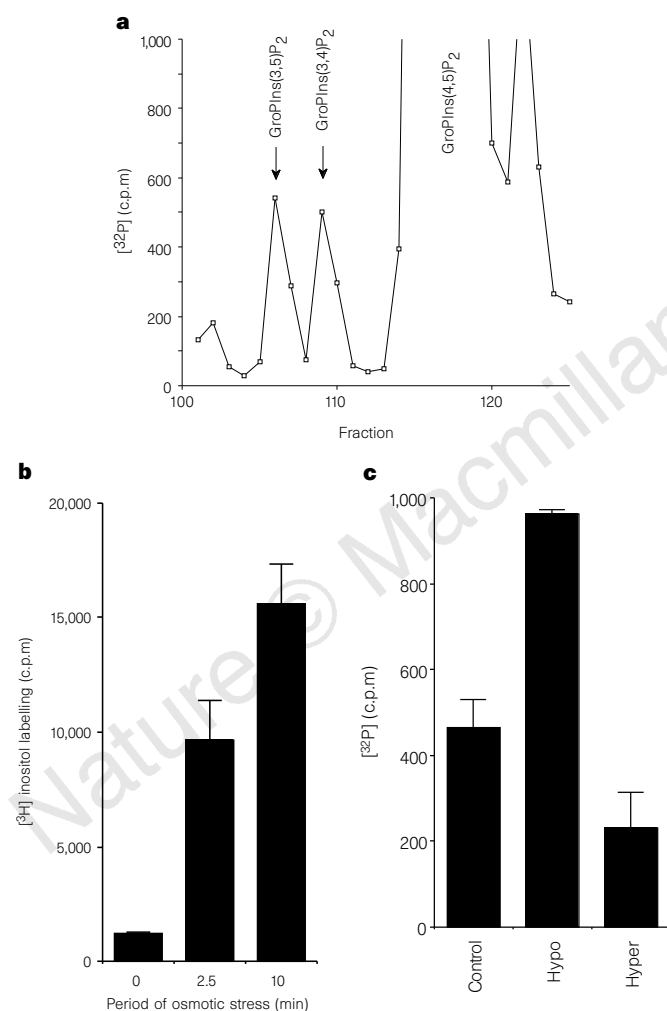


Figure 3 Demonstration of the presence of PtdIns(3,5)P₂ in *S. pombe* and in mammalian cells. **a**, HPLC chromatogram of deacylated lipids from ^{32}P -labelled Cos-7 cells, showing the presence of [^{32}P]PtdIns(3,5)P₂: the peak labelled GroPIns(3,5)P₂ co-chromatographed with [^3H]GroPIns(3,5)P₂ from *S. cerevisiae* (not shown; but see Fig. 1). **b**, Increased [^3H]PtdIns(3,5)P₂ content in [^3H]inositol-labelled *S. pombe* (wild type) hyperosmotically stressed with 0.9 M NaCl in YEPD medium for the periods defined. **c**, ^{32}P -labelling of PtdIns(3,5)P₂ in osmotically stressed Cos-7 cells labelled with [^{32}P]P_i and challenged for 10 min with unmodified medium (control), medium supplemented with 1 M sorbitol (Hypo) or medium diluted to quarter-strength with water (Hyper). As the recoveries of the stressed cells were variable, lipid labelling values were corrected to a reference value of 1.1×10^7 c.p.m. of total [^{32}P]lipid per dish. Data are representative of two independent experiments.

Table 1 ^{32}P in PtdIns(3,5)P₂ in hyperosmotically stressed yeast

Glycerophosphoester	^{32}P (d.p.m.)	^3H (d.p.m.)	$^{32}\text{P}:^3\text{H}$ ratio
GroPIns(3,5)P ₂	2,070	4,360	0.48
GroPIns5P	3,860	10,420	0.37
GroPIns	≤5	6,810	≤0.001
Calculated ^{32}P distribution: 5-phosphate, 78%; 3-phosphate, 22%; 1-phosphate, ≤0.2%			

Cells in phosphate-free medium were briefly labelled with [^{32}P]P_i (3 min), with 1.1 M NaCl present during the final 2 min, and the distribution of ^{32}P between the 1-, 3- and 5-phosphate groups of PtdIns(3,5)P₂ was determined (see Methods). Similar results were obtained in two experiments.

Table 2 Loss of hyperosmotic stimulation of PtdIns(3,5)P₂ synthesis in a *ts vps34* mutant

	PtdIns3P (d.p.m. $\times 10^{-3}$)	PtdIns(3,5)P ₂ (d.p.m. $\times 10^{-3}$)	PtdIns(4,5)P ₂ (d.p.m. $\times 10^{-3}$)
Parental strain (SEY6210)			
Permissive (30 °C): No shock	116 ± 2	3.9 ± 0.4	81 ± 3
10 min/0.9 M NaCl	110 ± 1	54 ± 3	136 ± 13
Restrictive (38 °C): No shock			
10 min/0.9 M NaCl	102 ± 1	1.9 ± 0.2	117 ± 0.8
10 min/0.9 M NaCl	121 ± 4	68 ± 6	219 ± 8
Temperature-sensitive in <i>vps34</i> (Phy102 + <i>pvps34</i> ^{tsf})			
Permissive (30 °C): No shock	121 ± 8	9.1 ± 0.7	101 ± 3
10 min/0.9 M NaCl	83 ± 5	65 ± 4	147 ± 38
Restrictive (38 °C): No shock			
10 min/0.9 M NaCl	21 ± 0.1	2.1 ± 0.5	107 ± 3
10 min/0.9 M NaCl	13 ± 0.6	1.9 ± 1.4	233 ± 29

The SEY6210 parental *S. cerevisiae* strain and a temperature-sensitive *vps34* mutant (Phy102 + *pvps34*^{tsf}) were labelled (see Methods) and then incubated at permissive or restrictive temperatures for 40 min. Cells were either immediately sampled or stressed in 0.9 M NaCl (see Methods) for 10 min, and their lipids analysed. The data presented (mean ± s.e.m.) are corrected to a constant cell number and are representative of results obtained in two independent experiments.

PtdIns(3,5)P₂ synthesis. We therefore compared the PtdIns3P content and PtdIns(3,5)P₂ accumulation response of normal cells and of a temperature-sensitive *vps34* mutant (PHY102 + *pvps34*^{tsf}). The PtdIns3P complement and PtdIns(3,5)P₂ accumulation response of the PHY102 + *pvps34*^{tsf} cells were normal at 30 °C, but at 38 °C these cells contained little PtdIns3P or PtdIns(3,5)P₂ and showed little or no hyperosmotic PtdIns(3,5)P₂ accumulation (Table 2). The parental cells showed a striking hyperosmotic activation of PtdIns(3,5)P₂ synthesis at both 30 °C and 38 °C. It therefore seems that PtdIns(3,5)P₂ synthesis, whether basal or osmotically stimulated, uses PtdIns3P that is synthesized by the *vps34*-encoded PtdIns 3-kinase as its substrate.

When we looked for PtdIns(3,5)P₂ in other eukaryotic cells, we obtained evidence to suggest that regulated PtdIns(3,5)P₂ synthesis may be more widespread in eukaryotes than either of the well defined PtdIns(4,5)P₂-based signalling pathways. An inositol lipid whose deacylation product migrated with GroPIns(3,5)P₂ was present in [³H]inositol-labelled *S. pombe*, in suspension-cultured plant cells ([³H]inositol-labelled *Daucus carota* cells: not shown) and in cultured mammalian cells (Fig. 3a; PtdIns(3,5)P₂ is also made from PtdIns3P in a fibroblast-like mouse cell line¹³). In *S. pombe*, hyperosmotic stress again increased the [³H]PtdIns(3,5)P₂ content (Fig. 3b). By contrast, hyperosmotic shock decreased [³²P]PtdIns(3,5)P₂ levels and hypo-osmotic shock enhanced PtdIns(3,5)P₂ labelling in monkey Cos-7 cells (Fig. 3c). Osmotic stresses have been reported to provoke changes in the total PtdInsP₂ levels (determined by TLC) of the alga *Dunaliella salina*¹⁹. Although these changes, which resemble those seen in yeast, were ascribed to PtdIns(4,5)P₂, they may involve PtdIns(3,5)P₂. Recent studies have often assumed that whenever a PtdInsP₂ isomer (or derived GroPInsP₂) can be separated from PtdIns(4,5)P₂ (or GroPIns(4,5)P₂), it is PtdIns(3,4)P₂. However, GroPIns(3,5)P₂ and GroPIns(3,4)P₂ behave similarly under many conditions, so recognition that PtdIns(3,5)P₂ is widespread makes it necessary to reassess at least some of this work on 'PtdIns(3,4)P₂'. In future, it will be necessary routinely to distinguish between GroPIns(3,5)P₂ and other PtdInsP₂ isomers.

The synthesis of PtdIns(3,5)P₂ by a PtdIns3P 5-kinase that is

quickly activated by hyperosmotic stress, independently of the *Hog1* pathway, suggests that PtdIns(3,5)P₂ plays a role in some other osmoregulatory pathway—maybe in an acute osmoprotective response. An important next step will be to identify the osmotically regulated PtdIns3P 5-kinase that is the pivotal enzyme in this previously unrecognized polyphosphoinositide-based signalling pathway, and to determine how it is activated. Many candidate phosphoinositide kinases have been identified, including the *S. cerevisiae* Tor1 and Tor2 proteins (and relatives such as Mei-41 (*Drosophila*), Rad3 (*S. pombe*), and FRAP (human)) and the products of two yeast PtdIns4P 5-kinase-like genes (*mss4* and *Fab1*)²⁰. The plethora of recently identified polyphosphoinositide-binding proteins, many of which contain pleckstrin-homology (PH) domains²¹, may include unrecognized downstream target proteins for PtdIns(3,5)P₂.

The *vps34*-encoded PtdIns 3-kinase supplies PtdIns3P for PtdIns(3,5)P₂ synthesis in *S. cerevisiae* (see above), so homologous type III 3-kinases in other cells^{7,18} probably do the same. Until now, the demonstration that *vps34* and related gene products are needed for protein trafficking, especially from Golgi to vacuole in yeast, has been the only major pointer to the function of the type III phosphoinositide 3-kinases. This led to an inference that PtdIns3P might play a direct role in such vesicle fluxes^{17,18}. Our results suggest the model summarized in Fig. 4. The type III 3-kinases may simply supply PtdIns3P for PtdIns(3,5)P₂ synthesis, and failure of PtdIns(3,5)P₂ synthesis may contribute to the characteristic defect in Golgi-to-vacuole trafficking in *vps34* cells. The fact that hyperosmotic stress provokes a rapid burst of PtdIns(3,5)P₂ synthesis might indicate that enhanced PtdIns(3,5)P₂-dependent vesicle trafficking is somehow involved in an acute osmotic adaptation response, which could involve acceleration or intensification of the normal Golgi-to-vacuole flux or the recruitment of some other vesicle trafficking event(s). □

Methods

[³H]inositol-labelled *S. cerevisiae*. For most experiments, *S. cerevisiae* (usually strains YHP1 or YPH499a) were grown almost to isotopic equilibrium (4–6 divisions) in an inositol-free and/or low-phosphate minimal (M) medium based on that described in ref. 22. Cells subcultured at 10⁵ cells ml⁻¹ in M medium containing 10 μCi ml⁻¹ [³H]inositol were grown at 30 °C to 1–2 × 10⁶ cells ml⁻¹, collected, twice washed in M medium, suspended in M medium and incubated at 30 °C for 15 min before an experiment. For some experiments (Figs 1b, 2b), cells were grown to mid-log phase in YEPD medium containing 50 μCi ml⁻¹ [³H]inositol (Amersham). After control samples had been removed, a stimulus (usually an equal volume of 1.8 M NaCl) was mixed with the cell suspension and samples were removed at intervals.

[³H]inositol-labelled yeast lipids: extraction and analysis. 100-μl aliquots of yeast suspension were mixed with 200 μl ice-cold CH₃OH:12 M HCl (100:1, v/v), vortexed for 10 s, frozen in liquid N₂ and thawed. Acid-washed glass beads (425–600 μm, 0.8 g; Sigma) were added, and the mixture was alternately vortexed for 1 min and cooled for 5 s with liquid N₂ (4 cycles): >90% of the yeast were disrupted. CH₃OH:12 M HCl (0.805 ml) and CHCl₃ (1.995 ml), 50 μl water and Folch inositol fraction (Sigma; 100 μg) were added. Phases were separated and the upper phase washed. The lipids (lower phase) were immediately deacylated (53 °C for 40 min) in monomethylamine reagent⁹, excess monomethylamine was removed, and glycerophosphoesters were recovered by partitioning into 1 ml of 50 mM HEPES/KOH against 2 ml butan-1-ol:petroleum ether (b.p. 40–60 °C):ethyl formate (20:4:1 by volume). The dried aqueous phases, redissolved and adjusted to pH 7–7.5, were stored at –20 °C. 100 ml *S. cerevisiae* YHP (1–2 × 10⁸ cells) typically yielded ~2 × 10⁸ d.p.m. of ³H-labelled alkali-labile lipids, of which 0.5–1% were PtdInsP₂ isomers. When necessary, glycerophosphoinositide esters were deglycerated⁹. HPLC on 250 × 4.0 mm Partisphere 5-SAX columns was used to analyse deacylated lipids²³ and InsP₃ molecules²⁴. Inositol monophosphates were separated by an acetate method²⁵.

Structural analysis of the novel GroPIns(4,5)P₂. *S. cerevisiae*, grown to ~4 × 10⁶ cells ml⁻¹ in medium containing 10 μCi ml⁻¹ [²⁻³H]inositol (NEN

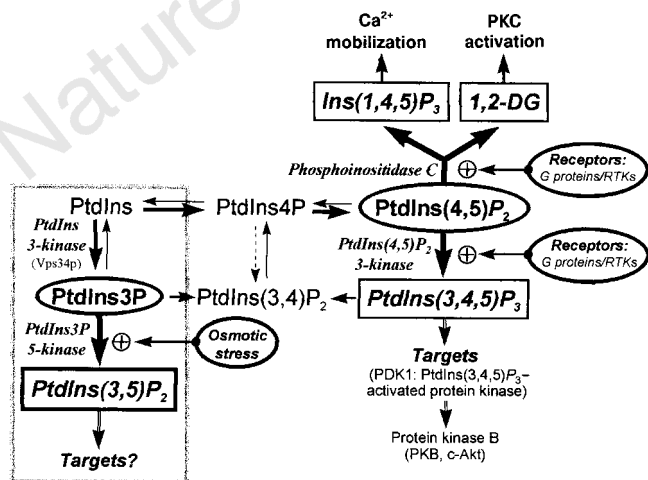


Figure 4 Scheme to illustrate the likely relationships between the stress-activated synthesis of PtdIns(3,5)P₂ described here (boxed, left) and previously known inositol lipid-based signalling mechanisms (1,2-DG, *sn*-1,2-diacylglycerol). Substrates of signalling reactions and their activators are enclosed in ovals, and proven or potential second messengers in rectangles. All of these reactions occur in animal cells, but the PtdIns(4,5)P₂ 3-kinase pathway appears to be absent from yeasts and plants, and the status of the phosphoinositidase C pathway is uncertain in yeasts.

or Amersham) were recovered and incubated at 30 °C for 15–30 min in 1.2 M sorbitol/10 mM potassium phosphate, pH 7.4/0.3% (v/v) β -mercaptoethanol and 50 ng ml⁻¹ Zymolase 20T (ICN). Spheroplast lipids were extracted (omitting the freezing and glass-bead steps) and deacylated. The unknown [³H]GroPInsP₂ was isolated by HPLC on a Partisphere 5-SAX anion-exchange column¹¹. Peak fractions were neutralised (20 μ l triethylamine), diluted and applied to a 0.3 ml Dowex AG-1 X8 column (formate form: BioRad), which was eluted with 0.3 M ammonium formate/0.1 formic acid (elutes P₁), water, and then 2 $\times$ 5 ml 1.0 M triethylamine bicarbonate (Fluka) (elutes [³H]GroPInsP₂). The final eluate (desalted [³H]GroPInsP₂) was dried *in vacuo* with a -90 °C trap, deglycerated⁹ and InsP₃ was isolated using an isocratic HPLC system²⁶. The resulting [³H]InsP₃ was mixed with [¹⁴C]Ins(3,4,5)P₂ (and/or [¹⁴C]Ins(1,5,6)P₃) from *Spirodela polyrrhiza* (a gift from C. Brearley) and treated with 0.1 M periodic acid (pH 3.0) at 30 °C for 5 d. Excess periodate was destroyed with NaBH₄, and the oxidized InsP₃ molecules were desphosphorylated. InsP₃-derived polyols were mixed with reference polyols (detected refractometrically), desalted and chromatographed on a water-eluted 300 $\times$ 7.8 mm Aminex HPX-87C column at 65 °C²⁶.

Analysis of the distribution of ³²P between the phosphate groups of PtdIns(3,5)P₂. *S. cerevisiae* (2–4 $\times$ 10⁹ cells) were washed twice, resuspended in phosphate-free medium and incubated at 30 °C for 10–15 min. 8–10 mCi of [³²P]P_i (PBS43, Amersham) were added, followed after 1 min by an equal volume of 2.2 M NaCl. Incubation was continued for 2 min, cells were killed by addition of ice-cold CH₃OH : H₂O : 12 M HCl (100 : 100 : 1, by volume), cooled in liquid N₂, and cell residues were sedimented and washed twice with CH₃OH : H₂O : 12 M HCl. ³²P-labelled lipids were extracted, resolved on an activated (120 °C for 3 h) 5 $\times$ 20 cm silica gel 60 TLC plate (Sigma) in CHCl₃ : CH₃OH : 15 M NH₄OH : H₂O (90 : 70 : 4 : 16, by volume) and detected by autoradiography. [³²P]PtdIns(3,5)P₂, which was completely resolved from [³²P]PtdIns(4,5)P₂, was isolated, deacylated, desalted and mixed with [³H]inositol-labelled GroPIns(3,5)P₂. The doubly labelled GroPIns(3,5)P₂ was either 3–4 mg dephosphorylated by incubation with washed erythrocyte membranes (3–4 mg protein) at pH 7.5 in 1 ml of 12.5 mM HEPES/KOH/1 mM EGTA/5 mM EDTA for 4 h¹⁰, or both monoester phosphate groups were removed with 400 units ml⁻¹ bovine intestinal alkaline phosphatase (Sigma) in 10 mM ethanolamine, pH 9.2 for 1 h.

Preparation of [¹⁴C]Ins1P and [¹⁴C]Ins2P. [¹⁴C]Ins1P was isolated by deacylation and deglyceration from the PtdIns of *S. cerevisiae* grown in [U-¹⁴C]inositol. A [¹⁴C]Ins1P sample treated with 1 M HCl (80 °C, 4 h) yielded a mixture of ¹⁴C-labelled DL-Ins1P, Ins2P and DL-Ins4P⁹.

Culture, labelling and stimulation of Cos-7 cells. Cells were grown in DMEM supplemented with 10% fetal bovine serum (Gibco) at 37 °C in 10% CO₂. They were washed once in phosphate-free DMEM and labelled for 1 h at 37 °C with 0.3 mCi ml⁻¹ [³²P]P_i (Amersham). Cells were washed in phosphate-free DMEM, appropriate media were added (Fig. 3c) and incubation was continued for 10 min. The medium was removed, cells were killed with 1 ml methanol/1M HCl (1 : 1), scraped off the plates and their lipids extracted and analysed.

Received 28 May; accepted 15 September 1997.

- Creba, J. A. *et al.* Rapid breakdown of phosphatidylinositol 4-phosphate and phosphatidylinositol 4,5-bisphosphate in rat hepatocytes stimulated by vasopressin and other Ca²⁺-mobilising hormones. *Biochem. J.* **212**, 733–747 (1983).
- Berridge, M. J. Inositol trisphosphate and diacylglycerol: two interacting second messengers. *Annu. Rev. Biochem.* **56**, 159–193 (1987).
- Berridge, M. J. Inositol trisphosphate and calcium signalling. *Nature* **361**, 315–325 (1993).
- Nishizuka, Y. The role of protein kinase C in cell surface signal transduction and tumour promotion. *Nature* **308**, 693–697 (1984).
- Nishizuka, Y. The molecular heterogeneity of protein kinase C and its implications for cellular regulation. *Nature* **334**, 661–665 (1990).
- Stephens, L., Jackson, T. R. & Hawkins, P. T. Agonist-stimulated synthesis of phosphatidylinositol (3,4,5)-trisphosphate: a new intracellular signalling system? *Biochim. Biophys. Acta* **1179**, 27–75 (1993).
- Toker, A. & Cantley, L. C. Signalling through the lipid products of phosphoinositide-3-OH kinase. *Nature* **387**, 673–676 (1997).
- De Camilli, P., Emr, S. D., McPherson, P. S. & Novick, P. Phosphoinositides as regulators in membrane traffic. *Science* **271**, 1533–1539 (1996).
- Letcher, A. J., Stephens, L. R. & Irvine, R. F. In *Methods in Inositol Research* (ed. Irvine, R. F.) 31–37 (Raven, London, 1990).
- Stephens, L. R., Hughes, K. T. & Irvine, R. F. Pathway of phosphatidylinositol (3,4,5)-trisphosphate synthesis in activated neutrophils. *Nature* **351**, 33–39 (1991).
- Auger, K. R., Serunian, L. A., Soltoff, S. P., Libby, P. & Cantley, L. C. PDGF-dependent tyrosine phosphorylation stimulates production of novel polyphosphoinositides in intact cells. *Cell* **57**, 167–175 (1989).

- Graziani, A., Ling, L. E., Endemann, G. & Cantley, L. C. Purification and characterisation of human erythrocyte phosphatidylinositol 4-kinase: phosphatidylinositol 4-kinase and phosphatidylinositol 3-phosphate 4-kinase are distinct enzymes. *Biochem. J.* **284**, 39–45 (1992).
- Whiteford, C. A., Brearley, C. A. & Ulug, E. T. Phosphatidylinositol 3,5-bisphosphate defines a new phosphatidylinositol 3-kinase pathway in resting mouse fibroblasts. *Biochem. J.* **323**, 597–601 (1997).
- Schuller, C., Brewster, J. L., Alexander, M., Gustin, M. & Ruis, H. The Hog pathway controls osmotic regulation of transcription via the stress-response element (stre) of the *Saccharomyces cerevisiae* *ctl1* gene. *EMBO J.* **13**, 4382–4389 (1994).
- Wurgler-Murphy, S. M. & Saito, H. Two-component signal transducers and MAPK cascades. *Trends Biochem. Sci.* **22**, 172–176 (1997).
- Doughney, C., McPherson, M. A. & Dormer, R. L. Metabolism of inositol 1,3,4,5-tetrakisphosphate by human erythrocyte membranes. *Biochem. J.* **251**, 927–929 (1988).
- Schu, P. V. *et al.* Phosphatidylinositol 3-kinase encoded by yeast *vps34* gene essential for protein sorting. *Science* **260**, 88–91 (1993).
- Vanhaesebroeck, B., Leeyers, S. J., Panayatou, G. & Waterfield, M. D. Phosphoinositide 3-kinases: a conserved family of signal transducers. *Trends Biochem. Sci.* **22**, 267–272 (1997).
- Einspahr, K. J., Peeler, R. C. & Thompson, G. A. Rapid changes in polyphosphoinositide metabolism associated with the response of *Dunaliella salina* to hypo-osmotic shock. *J. Biol. Chem.* **263**, 5775–5779 (1988).
- Yamamoto, A. *et al.* Novel PI(4)P 5-kinase homologue, Fab1p, essential for normal vacuole function and morphology in yeast. *Mol. Biol. Cell* **6**, 525–539 (1995).
- Divecha, N. & Irvine, R. F. Phospholipid signalling. *Cell* **80**, 269–278 (1995).
- Wickerham, L. J. A critical evaluation of the nitrogen assimilation tests commonly used in the classification of yeasts. *J. Bacteriol.* **52**, 293–301 (1946).
- Stephens, L. R., Hawkins, P. T. & Downes, C. P. Metabolic and structural evidence for the existence of a third species of polyphosphoinositide in cells, D-phosphatidyl-myo-inositol 3-phosphate. *Biochem. J.* **259**, 267–276 (1989).
- Wreggett, K. & Irvine, R. F. Automated isocratic hplc of inositol phosphate isomers. *Biochem. J.* **262**, 997–1000 (1989).
- Jenkinson, S. In *Methods in Molecular Biology: Signal Transduction Protocols* (eds Kendall, D. A. & Hill, S. J.) Vol. 41, 151–166 (Humana, Totowa, NJ, 1995).
- Barker, C. J. *et al.* Inositol 1,2,3-trisphosphate and inositol 1,2- and/or 2,3-bisphosphate are normal constituents of mammalian cells. *Biochem. J.* **306**, 557–564 (1995).
- Woscholski, R., Waterfield, M. D. & Parker, P. J. Purification and biochemical characterization of a mammalian phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase. *J. Biol. Chem.* **270**, 31001–31007 (1995).

Acknowledgements. We thank the Royal Society (R.H.M. and L.R.S.), the MRC (R.H.M.), and the Beit Memorial Trust (S.K.D.) for financial support. We also thank S. Emr and M. J. Gustin for yeast strains and for discussion.

Correspondence and requests for materials should be addressed to S.K.D. (e-mail: s.k.dove@bham.ac.uk).

A new pathway for synthesis of phosphatidylinositol-4,5-bisphosphate

Lucia E. Rameh*, Kimberley F. Tolias, Brian C. Duckworth & Lewis C. Cantley

Department of Cell Biology, Harvard Medical School, and Division of Signal Transduction, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA

Phosphatidylinositol-4,5-bisphosphate (PtdIns-4,5-P₂), a key molecule in the phosphoinositide signalling pathway, was thought to be synthesized exclusively by phosphorylation of PtdIns-4-P at the D-5 position of the inositol ring. The enzymes that produce PtdIns-4,5-P₂ *in vitro* fall into two related subfamilies (type I and type II PtdInsP-5-OH kinases, or PIP(5)Ks) based on their enzymatic properties and sequence similarities¹. Here we have reinvestigated the substrate specificities of these enzymes. As expected, the type I enzyme phosphorylates PtdIns-4-P at the D-5 position of the inositol ring. Surprisingly, the type II enzyme, which is abundant in some tissues, phosphorylates PtdIns-5-P at the D-4 position, and thus should be considered as a 4-OH kinase, or PIP(4)K. The earlier error in characterizing the activity of the type II enzyme is due to the presence of contaminating PtdIns-5-P in commercial preparations of PtdIns-4-P. Although PtdIns-5-P was previously thought not to exist *in vivo*, we find evidence for the presence of this lipid in mammalian fibroblasts, establishing a new pathway for PtdIns-4,5-P₂ synthesis.

While characterizing the enzymatic activity of PtdIns-4,5-P₂ phosphatases, we discovered that type I and type II phosphatidylinositol-4-phosphate 5-hydroxy kinases (PIP(5)Ks) phosphorylated different sites on the inositol ring. Using commercial PtdIns-4-P as

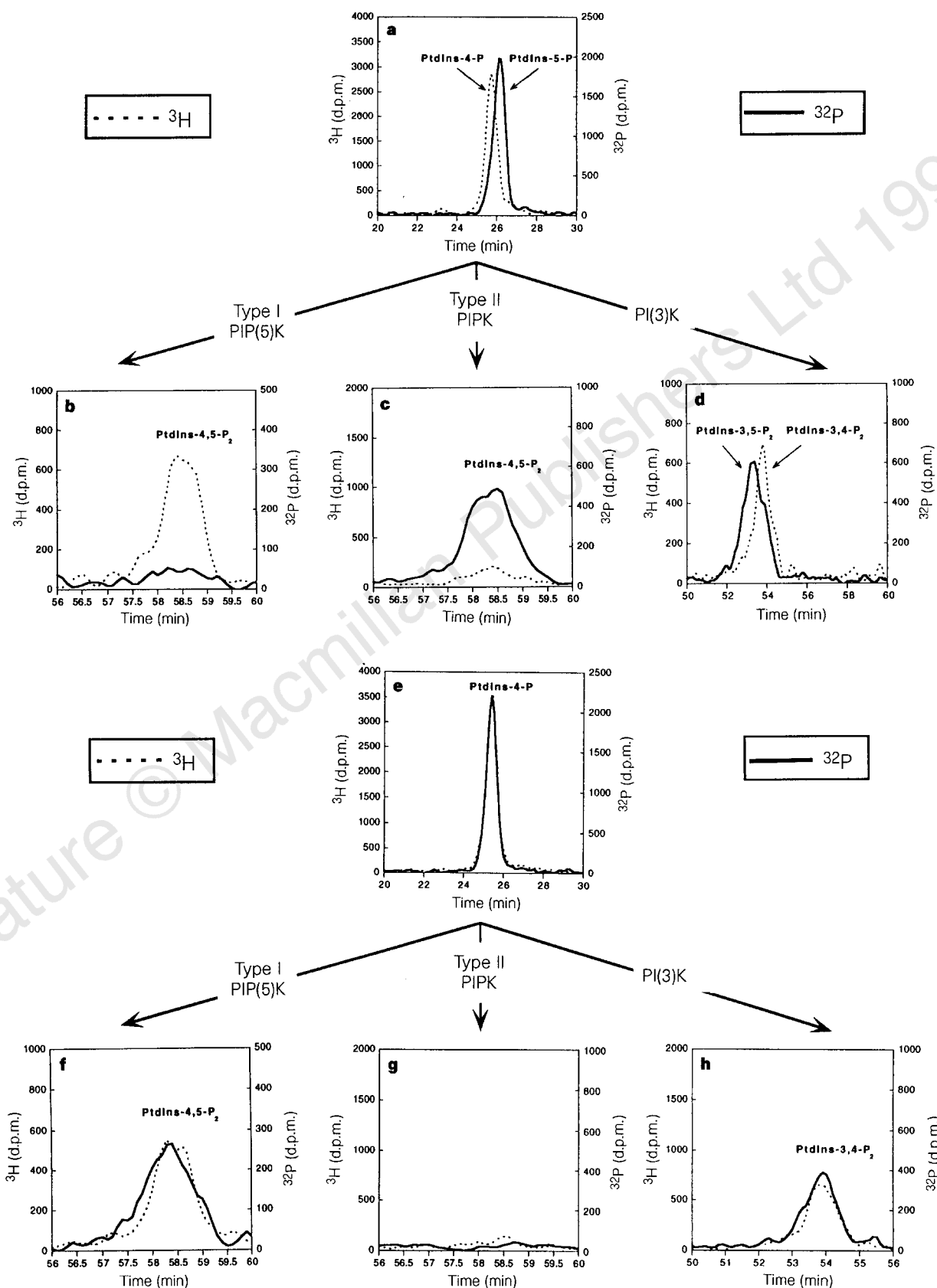


Figure 1 Type II PIPK phosphorylates PtdIns-5-P. ^{32}P -labelled PtdIns-4,5- P_2 , generated by phosphorylation of commercial PtdIns-4-P with type I PIPK (**a**; solid line) was dephosphorylated by SHIP. The PtdIns-P product was purified by TLC and used as a substrate for type I PIPK (**b**), type II PIPK (**c**) or PI(3)K (**d**), using unlabelled ATP. Similarly, ^{32}P -PtdIns-4,5- P_2 , generated by phosphorylation of commercial PtdIns-4-P with type II PIPK, was dephosphorylated by SHIP. The PtdIns-P product (**e**; solid line) was purified by TLC and used as a substrate for the same three enzymes (**f**, **g** and **h**). ^3H -labelled PtdIns-4-P (dotted lines) was added to each reaction as an internal control. Lipids were deacylated and analysed by HPLC. Peaks corresponding to the deacylated form of each phosphoinositide are indicated. Results are representative of more than 3 experiments.

substrate and [^{32}P]ATP, we synthesized [^{32}P]PtdIns-4,5- P_2 with either type I or type II PIP(5)K. Products were separately treated with the SH2-domain-containing inositol phosphatase (SHIP), which preferentially dephosphorylates position 5 of the inositol ring of PtdIns-3,4,5- P_3 (ref. 2). [^3H]PtdIns-4,5- P_2 was added to each reaction mixture as an independent measure of PtdIns-4,5- P_2 hydrolysis. Both [^3H]- and [^{32}P]-labelled PtdIns-4,5- P_2 were hydrolysed, and when [^{32}P]-PtdIns-4,5- P_2 produced by the type I enzyme was used as substrate, most of the [^{32}P] was released as inorganic phosphate, consistent with a preference of this phosphatase for position 5 of the inositol ring (results not shown). However, a small amount ($\sim 5\%$) of [^{32}P]PtdIns-5-P was produced. When analysed by high-pressure liquid chromatography (HPLC), deacylated [^{32}P]PtdIns-5-P migrated ~ 0.5 min later than deacylated [^3H]PtdIns-4-P derived from hydrolysis of the [^3H]PtdIns-4,5- P_2 standard in the same mixture (Fig. 1a). The [^3H]-labelled product revealed a small shoulder at the position of PtdIns-5-P. Deacylated PtdIns-5-P migrated 0.5 min later on HPLC than a deacylated [^{32}P]PtdIns-4-P standard made by using phosphatidylinositol 4-OH kinase- β (ref. 3, and data not shown). The relative migratory positions of PtdIns-4-P and PtdIns-5-P are consistent with previous observations⁴. Thus, in agreement with previous findings, type I PIP(5)K preferentially phosphorylates position 5 of PtdIns-4-P and SHIP preferentially dephosphorylates position 5, although there is some dephosphorylation at position 4 when PtdIns-4,5- P_2 is used as substrate (L.E.R. *et al.*, manuscript in preparation).

Surprisingly, SHIP-dependent dephosphorylation of [^{32}P]PtdIns-4,5- P_2 produced by type II PIP(5)K phosphorylation of commercial PtdIns-P resulted in almost quantitative recovery of [^{32}P]-labelled PtdIns-4-P (for example, the [^{32}P]/ ^3H ratio in the substrate and product were similar, indicating that dephosphorylation of position 5 did not cause significant loss of radioactivity from the lipid) (results not shown). This indicated that the type II enzyme phosphorylates the D-4 position. Deacylated [^{32}P]PtdIns-P produced by SHIP co-migrates on HPLC with [^3H]PtdIns-4-P produced in the same reaction, confirming the identity of PtdIns-4-P as the product (Fig. 1e). This product also co-migrated with [^{32}P]PtdIns-4-P standard made by phosphatidylinositol 4-OH kinase- β (ref. 3, and data not shown). The ability of the type II enzyme to phosphorylate the D-4 position and produce a molecule that exactly co-migrates with PtdIns-4,5- P_2 when PtdIns-4-P is provided as substrate suggests that commercial PtdIns-4-P also contains PtdIns-5-P. The alternative explanation that this enzyme phosphorylates both the 4 and 5 positions of contaminating PtdIns is unlikely as this enzyme does not phosphorylate pure PtdIns^{5,6}.

To confirm that the type II enzyme preferentially phosphorylates PtdIns-5-P, we compared the relative ability of this enzyme to phosphorylate enzymatically produced [^{32}P]PtdIns-5-P and [^{32}P]PtdIns-4-P using unlabelled ATP; [^3H]PtdIns-4-P was added to each substrate as an internal standard. We found that the type II enzyme converted [^{32}P]PtdIns-5-P to [^{32}P]PtdIns-4,5- P_2 (38% conversion under these conditions) but produced virtually no

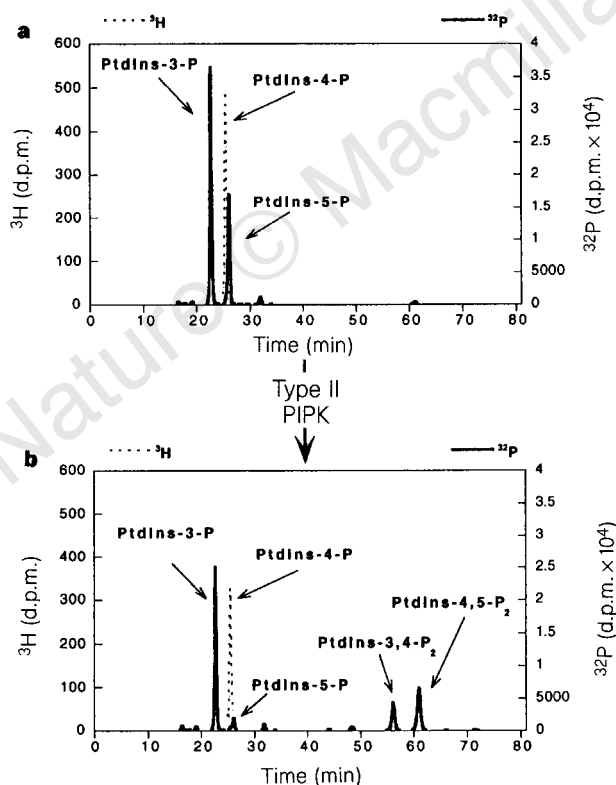


Figure 2 PtdIns-5-P is the preferred substrate for the type II PIPK. A mixture containing [^{32}P]PtdIns-3-P, [^{32}P]PtdIns-5-P and [^3H]PtdIns-4-P was incubated without (a) or with (b) type II PIPK for 2 h at 37°C. Lipids were chloroform-extracted, deacylated and analysed by HPLC. Dotted lines represent measurements from the ^3H channel; solid lines represent measurements from the ^{32}P channel.

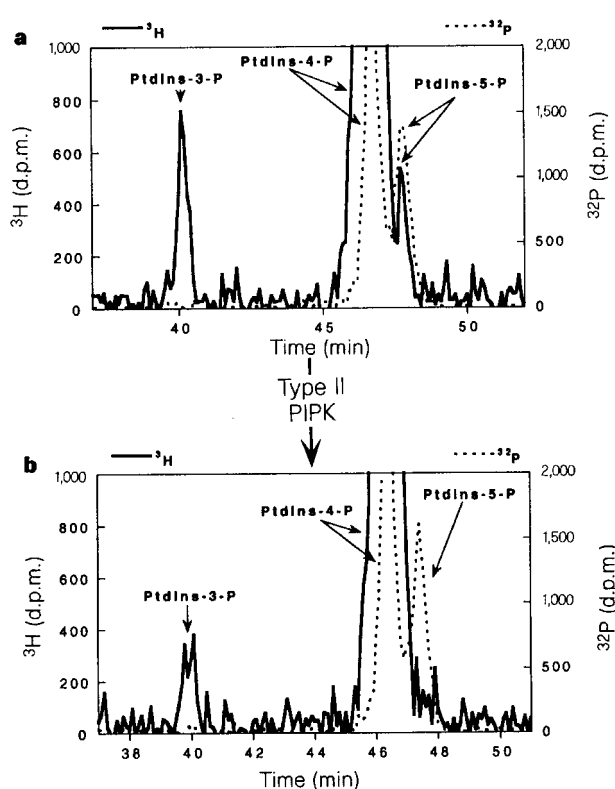


Figure 3 PtdIns-5-P is found *in vivo*. Exponentially growing NIH3T3 cells were labelled *in vivo* with ^3H -inositol. Cells were lysed, lipids were extracted and redissolved in aqueous buffer, sonicated and incubated without (a) or with (b) type II PIPK for 2 h at 37°C with unlabelled ATP. ^3H -labelled lipids (solid lines) were re-extracted, deacylated and analysed by HPLC. Deacylated [^{32}P]-labelled PtdIns-4-P and [^{32}P]-labelled PtdIns-5-P were used as internal standards (dotted lines). Peaks corresponding to the deacylated form of each phosphoinositide are indicated. Result is representative of two experiments.

product when [32 P]PtdIns-4-P was used as substrate (compare Fig. 1c and g). The deacylated product co-migrated with a deacylated [3 H]PtdIns-4,5- P_2 standard and with deacylated [32 P]PtdIns-4,5- P_2 produced by type I PIP(5)K (see later). In both reactions, a very small amount of [3 H]PtdIns-4,5- P_2 was produced, confirming that [3 H]PtdIns-4-P in the same reaction mixture is also a poor substrate for type II PIP(5)K. Thus, the type II enzyme preferentially uses PtdIns-5-P as a substrate and produces PtdIns-4,5- P_2 .

To confirm that type I and type II enzymes carry out different reactions, the same two substrates were used to investigate the specificity of type I PIP(5)K. This enzyme converted both [32 P]PtdIns-4-P and [3 H]PtdIns-4-P to PtdIns-4,5- P_2 (Fig. 1f), but converted virtually none of the [32 P]PtdIns-5-P to [32 P]PtdIns-4,5- P_2 (Fig. 1b), demonstrating that both enzymes can produce the same product (PtdIns-4,5- P_2) but by phosphorylating different substrates.

The difference in the [32 P]PtdIns-4-P and [32 P]PtdIns-5-P substrates was further shown by phosphorylating these lipids with recombinant phosphoinositide 3-OH kinase (PI(3)K). PI(3)K converted [32 P]PtdIns-4-P and [3 H]PtdIns-4-P to PtdIns-3,4- P_2 (Fig. 1h), in agreement with previous results⁷. In contrast, PI(3)K converted PtdIns-5-P to a molecule whose deacylation product migrated ahead of deacylated PtdIns-3,4- P_2 (Fig. 1d), at the position expected for deacylated PtdIns-3,5- P_2 (refs 4, 7, 8). These results are in agreement with our identification of PtdIns-4-P and PtdIns-5-P.

We also tested whether the type II enzyme could phosphorylate PtdIns-3-P. In a mixture of [32 P]PtdIns-5-P and [32 P]PtdIns-3-P (of similar specific radioactivity), the type II enzyme preferentially phosphorylated [32 P]PtdIns-5-P (60–80% conversion; Fig. 2). Some [32 P]PtdIns-3,4- P_2 was produced (5–20% conversion), indicating that PtdIns-3-P is a substrate for this enzyme. The product was identified as PtdIns-3,4- P_2 on the basis of its co-migration with [3 H]PtdIns-3,4- P_2 , produced by dephosphorylating [3 H]PtdIns-3,4,5- P_3 at the 5 position using SHIP, and its resolution from a PtdIns-3,5- P_2 standard (not shown). No [32 P]PtdIns-3,4,5- P_3 was generated by type II PIPK (results not shown). Thus, the type II enzyme phosphorylates the 4 position of both PtdIns-5-P and PtdIns-3-P but prefers the former.

Synthetic dipalmitoyl PtdIns-5-P and dipalmitoyl PtdIns-3-P were also used as substrates for type II PIPK. Synthetic dipalmitoyl PtdIns-5-P was converted to PtdIns-4,5- P_2 by the type II enzyme but not by type I PIP(5)K (not shown), in agreement with the results shown in Fig. 1. When 40 μ M synthetic dipalmitoyl PtdIns-5-P and 40 μ M dipalmitoyl PtdIns-3-P were simultaneously used as substrates for the type II enzyme, PtdIns-4,5- P_2 was the main product (not shown), in agreement with the results shown in Fig. 2.

Finally, we investigated whether PtdIns-5-P exists in mammalian cells by growing NIH3T3 fibroblasts in the presence of 3 H-inositol. The labelled lipids were extracted, deacylated and analysed by HPLC. Using a shallow ammonium phosphate gradient, [32 P]PtdIns-4-P eluted one minute before [32 P]PtdIns-5-P (Fig. 3, dotted lines). Next to the [3 H]PtdIns-4-P peak and co-migrating with the [32 P]PtdIns-5-P marker was a small peak that was 2% as high as the PtdIns-4-P peak (Fig. 3a, solid line). The poor separation of PtdIns-5-P from PtdIns-4-P on conventional HPLC gradients and the excess PtdIns-4-P in cells could explain why PtdIns-5-P had been missed previously. To confirm that this peak was PtdIns-5-P, the *in vivo* labelled 3 H-phosphoinositides from NIH3T3 cells were treated with type II PIPK and unlabelled ATP, deacylated and analysed by HPLC: virtually all of the radioactivity in the peak corresponding to [3 H]PtdIns-4-P was recovered as PtdIns-4-P after enzyme treatment and the 3 H-labelled peak that co-migrated with [32 P]PtdIns-5-P was no longer evident, indicating that this phosphoinositide was a substrate for type II PIPK. Some of the peak corresponding to [3 H]PtdIns-3-P was also lost after this treatment, as predicted from the results shown in Fig. 2. These

experiments indicate that PtdIns-5-P is present in exponentially growing fibroblasts in amounts comparable to PtdIns-3-P (2 and 5%, respectively).

We have found that type II PIP(5)K, the first PtdInsP kinase to be purified and cloned^{5,9,10}, is actually a PtdIns-5-P 4-kinase rather than a PtdIns-4-P 5-kinase, as previously assumed, so this enzyme should be renamed PIP(4)K. We have also shown, to our knowledge for the first time, that PtdIns-5-P, the preferred substrate for this enzyme, is present in mammalian cells, at levels comparable to PtdIns-3-P. The existence of large amounts of the type II enzyme in various tissues, particularly brain, red cells and platelets^{5,9–11}, indicates that production of PtdIns-4,5- P_2 by this alternative pathway may be important in several types of cell.

The previous assumption that all PtdIns-4,5- P_2 is produced by the canonical pathway (via a PtdIns-4-P intermediate) was based on the following evidence: no PtdIns-5-P had been found *in vivo*; no PtdIns-5-P 4-kinase activity had been detected; [32 P]PtdIns-4,5- P_2 produced by brief labelling of several cell types with 32 PO $_4^{3-}$ had more radioactivity in the 5 position than in the 4 position, indicating that the 5 phosphate is added last⁴. Our results demonstrate the existence of the enzyme and substrate needed to make PtdIns-4,5- P_2 by the alternative pathway. The pulse-labelling experiments⁴ do not preclude the alternative pathway: they indicate that most PtdIns-4,5- P_2 is produced via a PtdIns-4-P intermediate under those conditions. The alternative pathway will need to be put in context by using gene-knockout and dominant-negative experiments. Our results indicate that synthesis of PtdIns-5-P will be limiting in the alternative pathway.

Members of the type II PIPK family are associated with or regulated by various signal-transduction pathways. For example, the platelet type II enzyme associates with the cytoskeleton and undergoes a change in phosphorylation status in response to thrombin¹¹, and a member of the type II family associates with the TNF receptor and appears to be regulated by the cytokine TNF- α ¹². PtdIns-4,5- P_2 is not only a precursor of Ins-1,4,5- P_3 and PtdIns-3,4,5- P_3 , but is also a regulator of the actin cytoskeleton and of vesicle trafficking¹. Regulation of the synthesis of this lipid is evidently more complicated than previously assumed and needs reinvestigation.

A recent report has indicated that both type I PIP(5)K and type II PIPK can phosphorylate PtdIns-3-P to PtdIns-3,4- P_2 (ref. 13). We have also shown that PtdIns-3-P can be phosphorylated at D-4 by the type II enzyme (Fig. 2), but we find that type I PIP(5)K converts PtdIns-3-P to PtdIns-3,5- P_2 (K.T. *et al.*, manuscript in preparation). Although PtdIns-3-P is a poor substrate for type II PIPK compared with PtdIns-5-P, PtdIns-3-P may be the principal substrate of this enzyme in cells lacking PtdIns-5-P. As proposed by Zhang *et al.*¹³, the type II enzyme could account for the PtdIns-3-P 4-OH kinase activity in red cells and platelets^{14–16}. □

Methods

Lipid kinase and phosphatase reactions. For lipid kinase reactions, we used the indicated lipid substrate (Sigma) and recombinant murine type I PIPK- β ¹⁷, red-cell purified type II PIPK³, or recombinant type II PIPK- α ¹⁰, as described³. Deacylated lipids were analysed by anion-exchange high-pressure liquid chromatography (HPLC) using a Partisphere SAX column (Whatman), as described¹⁸. Synthetic dipalmitoyl PtdIns-3-P and dipalmitoyl PtdIns-5-P were provided by G. Prestwich. Dephosphorylation of PtdIns-4,5- P_2 by recombinant SHIP was performed as described².

***In vivo* labelling of lipids.** NIH3T3 (ATCC) cells were labelled at 37°C with [3 H]inositol (4 μ Ci ml⁻¹) (ARC) for 24 h in inositol-free Dulbecco's modified Eagle medium supplemented with 10% dialysed fetal calf serum. The medium was then replaced with complete DMEM-FCS for 1 h. Lipids were extracted as described¹⁸ and resuspended in aqueous buffer, sonicated and incubated in the presence or absence of type II PIPK and ATP. After 1 h at 37°C, the lipids were chloroform-extracted and deacylated as described. Deacylated lipids were analysed by HPLC using a shallow ammonium phosphate gradient (10–

35 mM ammonium phosphate, pH 3.8, over 50 min) to separate PtdIns-4-P from PtdIns-5-P.

Received 20 August; accepted 23 September 1997.

- Loijens, J. C., Boronenkov, I. V., Parker, G. J. & Anderson, R. A. The phosphatidylinositol 4-phosphate 5-kinase family. *Adv. Enz. Reg.* **36**, 115–140 (1996).
- Damen, J. E. *et al.* The 145-kDa protein induced to associate with Shc by multiple cytokines is an inositol tetraphosphate and phosphatidylinositol 3,4,5-triphosphate 5-phosphatase. *Proc. Natl. Acad. Sci. USA* **93**, 1689–1693 (1996).
- Meyers, R. & Cantley, L. C. Cloning and characterization of a wortmannin-sensitive human phosphatidylinositol 4-kinase. *J. Biol. Chem.* **272**, 4385–4390 (1997).
- Whiteford, C. C., Brearley, C. A. & Ulug, E. T. Phosphatidylinositol 3,5-bisphosphate defines a novel PI 3-kinase pathway in resting mouse fibroblasts. *Biochem. J.* **323**, 597–601 (1997).
- Ling, L. E., Schulz, J. T. & Cantley, L. C. Characterization and purification of membrane-associated phosphatidylinositol-4-phosphate kinase from human red blood cells. *J. Biol. Chem.* **264**, 5080–5088 (1989).
- Bazenet, C. E., Ruano, A. R., Brockman, J. L. & Anderson, R. A. The human erythrocyte contains two forms of phosphatidylinositol-4-phosphate 5-kinase which are differentially active toward membranes. *J. Biol. Chem.* **265**, 18012–18022 (1990).
- Auger, K. R., Serunian, L. A., Soltoff, S. P., Libby, P. & Cantley, L. C. PDGF-dependent tyrosine phosphorylation stimulates production of novel polyphosphoinositides in intact cells. *Cell* **57**, 167–175 (1989).
- Carpenter, C. L. & Cantley, L. C. Phosphoinositide kinases. *Biochemistry* **29**, 11147–11156 (1990).
- Boronenkov, I. V. & Anderson, R. A. The sequence of phosphatidylinositol-4-phosphate 5-kinase defines a novel family of lipid kinases. *J. Biol. Chem.* **270**, 2881–2884 (1995).
- Divecha, N., Truong, O., Hsuan, J. J., Hinchcliffe, K. A. & Irvine, R. F. The cloning and sequence of the C isoform of PtdIns4P 5-kinase. *Biochem. J.* **309**, 715–719 (1995).
- Hinchcliffe, K. A., Irvine, R. F. & Divecha, N. Aggregation-dependent, integrin-mediated increases in cytoskeletonally associated PtdInsP₂ (4,5) levels in human platelets are controlled by translocation of PtdIns 4-P 5-kinase C to the cytoskeleton. *EMBO J.* **15**, 6516–6524 (1996).
- Castellino, A. M., Parker, G. J., Boronenkov, I. V., Anderson, R. A. & Chao, M. V. A novel interaction between the juxtamembrane region of the p55 tumor necrosis factor receptor and phosphatidylinositol-4-phosphate 5-kinase. *J. Biol. Chem.* **272**, 5861–5870 (1997).
- Zhang, X. *et al.* Phosphatidylinositol-4-phosphate 5-kinase isozymes catalyze the synthesis of 3-phosphate-containing phosphatidylinositol signaling molecules. *J. Biol. Chem.* **272**, 17756–17761 (1997).
- Yamamoto, K., Graziani, A., Carpenter, C., Cantley, L. C. & Lapetina, E. G. A novel pathway for the formation of phosphatidylinositol 3,4-bisphosphate. Phosphorylation of phosphatidylinositol 3-monophosphate by phosphatidylinositol-3-monophosphate 4-kinase. *J. Biol. Chem.* **265**, 22086–22089 (1990).
- Yamamoto, K. & Lapetina, E. G. Protein kinase C-mediated formation of phosphatidylinositol 3,4-bisphosphate in human platelets. *Biochem. Biophys. Res. Commun.* **168**, 466–472 (1990).
- Graziani, A., Ling, L. E., Endemann, G., Carpenter, C. L. & Cantley, L. C. Purification and characterization of human erythrocyte phosphatidylinositol 4-kinase. Phosphatidylinositol 4-kinase and phosphatidylinositol 3-monophosphate 4-kinase are distinct enzymes. *Biochem. J.* **284**, 39–45 (1992).
- Ishihara, H. *et al.* Cloning of cDNAs encoding two isoforms of 68-kDa type I phosphatidylinositol-4-phosphate 5-kinase. *J. Biol. Chem.* **271**, 23611–23614 (1996).
- Serunian, L. A., Auger, K. R. & Cantley, L. C. Identification and quantification of polyphosphoinositides produced in response to platelet-derived growth factor stimulation. *Meth. Enzymol.* **198**, 78–87 (1991).

Acknowledgements. We thank G. Preswich for synthetic PtdIns-5-P, K. Hinchcliffe for recombinant type II PIPK, H. Ishihara for type I PIPK cDNA, N. Ravinchandran for the GST-SHIP expression vector, R. Meyers for recombinant PI-4K β , A. Couvillon for preparing recombinant type I PIPK and recombinant PI-3K, and D. Fruman and C. Carpenter for critically reading this manuscript. L.E.R. is supported by The Medical Foundation–Charles King Trust. This research was supported by the NIH.

Correspondence and requests for materials should be addressed to L.E.R. (e-mail: lrameh@bidmc.harvard.edu).

Folding dynamics and mechanism of β -hairpin formation

Victor Muñoz, Peggy A. Thompson, James Hofrichter & William A. Eaton

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA

Protein chains coil into α -helices and β -sheet structures. Knowing the timescales and mechanism of formation of these basic structural elements is essential for understanding how proteins fold¹. For the past 40 years, α -helix formation has been extensively investigated in synthetic and natural peptides^{2–5}, including by nanosecond kinetic studies^{6,7}. In contrast, the mechanism of formation of β structures has not been studied experimentally. The minimal β -structure element is the β -hairpin, which is also the basic component of antiparallel β -sheets. Here we use a

nanosecond laser temperature-jump apparatus to study the kinetics of folding a β -hairpin consisting of 16 amino-acid residues. Folding of the hairpin occurs in 6 μ s at room temperature, which is about 30 times slower than the rate of α -helix formation^{6,7}. We have developed a simple statistical mechanical model that provides a structural explanation for this result. Our analysis also shows that folding of a β -hairpin captures much of the basic physics of protein folding, including stabilization by hydrogen bonding and hydrophobic interactions, two-state behaviour, and a funnel-like, partially rugged energy landscape.

Finding protein fragments or designed peptides for the investigation of β -hairpin formation has been difficult because of instability and/or aggregation. However, NMR has revealed a significant population of β -hairpin structure in several monomeric peptides^{8–14}. We have studied one such peptide^{9,10} (Fig. 1). The sequence contains a single tryptophan (W43), which provides an intrinsic probe for both equilibrium and kinetic experiments. Formation of the β -hairpin is accompanied by an increase in W43 fluorescence because of its consolidation into a hydrophobic cluster with phenylalanine 52 (F52) and valine 54 (V54). By adding a dansylated lysine to the carboxy terminus of the peptide, we have used tryptophan $\rightarrow$ dansyl excitation energy transfer as an additional probe for the degree of β -hairpin structure. Equilibrium measurements with both methods have given consistent results (Fig. 2 and Methods). Using a two-state analysis, in which the peptide is either unfolded (W43 exposed to the solvent) or folded (W43 in the β -hairpin hydrophobic cluster), we obtain apparent thermodynamic parameters for the folding transition: $\Delta H = -11.6$ kcal mol⁻¹, $\Delta S = -39$ cal mol⁻¹ K⁻¹. The population of the β -hairpin obtained at the lowest temperature studied (273 K) is $\sim 80\%$ (inset to Fig. 2), in reasonable agreement with the NMR determination of the β -hairpin population¹⁰.

To study the β -hairpin kinetics we monitored the tryptophan fluorescence following nanosecond laser temperature jumps of ~ 15 degrees, with final temperatures in the range 288–328 K. After completion of a temperature jump from 273 to 288 K, the tryptophan fluorescence decreases in a single exponential relaxation to its new equilibrium value in which the hairpin population has decreased by $\sim 15\%$ (Fig. 3a). The time constant for this relaxation is 3.7 ± 0.3 μ s (s.d. from 5 experiments). In the dansylated peptide, a single relaxation with the same rate is observed ($\tau = 3.5 \pm 0.5$ μ s) (Fig. 3b). In this case, however, the fluorescence increases because

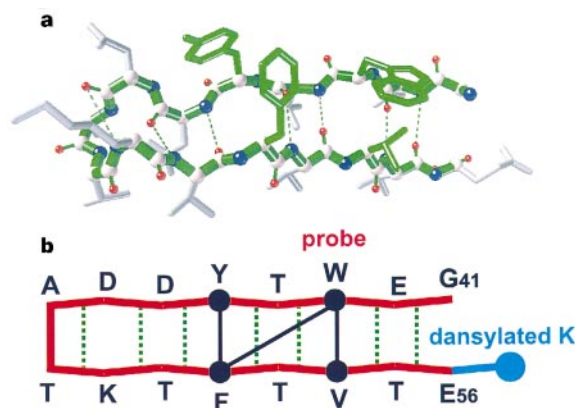


Figure 1 The β -hairpin from protein G B1. **a**, Three-dimensional structure of the C-terminal fragment (41–56) of protein G B1 in the intact protein (1GB1.PDB)²⁴ with the four residues of the hydrophobic cluster shown in green. **b**, Schematic representation of the structure showing the pattern of hydrogen (green) bonds and the three (blue) interactions between hydrophobic side chains that stabilize the β -hairpin structure. Observation of NOEs between the side chains of residue pairs W43-F52, W43-V54, Y45-F52 in the fragment (41–56) of protein G B1 indicate that the hydrophobic cluster is preserved in the isolated structure¹⁰.

the average tryptophan–dansyl distance is larger in the unfolded molecules, resulting in less quenching by Forster excitation energy transfer. The agreement in the relaxation times measured by both methods further reinforces our interpretation that the fluorescence monitors the hairpin population. It also suggests that there is a single effective thermodynamic barrier separating the folded and unfolded peptide, producing a two-state kinetic system. Assuming perfect two-state behaviour, the β -hairpin kinetics are simply described by a folding rate (with a negative activation energy) and an unfolding rate that are directly calculated from the equilibrium constants and relaxation rates at different temperatures. At the midpoint of the unfolding transition at 297 K, the folding and unfolding rates for the β -hairpin are both equal to $1/(6 \mu\text{s})$. The rate of formation of this β -hairpin is, therefore, much slower than the rate of $1/(180 \text{ ns})$ recently measured for α -helix formation in a polyaniline-based helical peptide^{6,7}.

To explain the two-state behaviour and slower formation rate for the β -hairpin, we have developed a statistical mechanical model which grasps the structural complexity of the hairpin with a minimal number of physically meaningful parameters. The model assumes that the stability of this hairpin can be described by only three thermodynamic factors: (1) loss of conformational entropy upon fixing the peptide bonds in their native conformation (ΔS_{conf}) which opposes β -hairpin formation; (2) stabilization by backbone hydrogen bonds between the two strands (ΔH_{hb}); and (3) stabilization by the three side-chain interactions in the hydrophobic cluster (ΔG_{sc}) (Fig. 1b), which for simplicity we assume to be independent of temperature. Two possible structural states are defined for each of the 15 peptide bonds: native (native ψ of residue i , and native ϕ of residue $i + 1$) and non-native (all other ψ , ϕ pairs). The number of molecular species in the model is reduced from 2^{15} ($= 32,768$) to 121 by assuming that there is no more than one stretch of contiguous native peptide bonds in each molecule (the single sequence approximation). The hairpin structure therefore 'grows' by adding native peptide bonds adjacent to already existing native peptide bonds. By postulating a transition state for elementary kinetic steps in which the new peptide bond already has the native conformation but no stabilizing interactions, only two additional parameters are required for a complete kinetic description of the system, an activation energy (E_0) and a pre-exponential factor (k_0). The partition function and system of differential equations describing the species population as a function of time can be readily evaluated (V.M. *et al.*, manuscript in preparation). Using the kinetic and thermodynamic constants that reproduce the

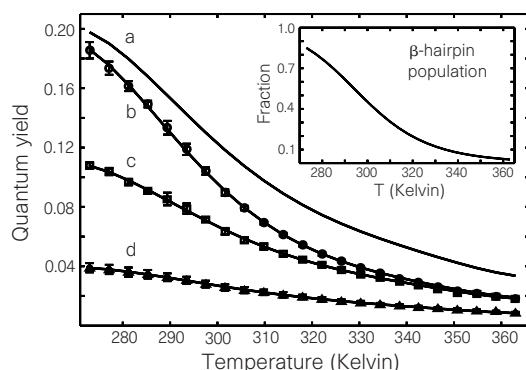


Figure 2 Thermal unfolding of the β -hairpin measured by tryptophan fluorescence (pH 7, 20 mM potassium phosphate). Curve a, calculated fluorescence of the folded hairpin. Curve b, circles, β -hairpin peptide. Curve c, squares, reference peptide GEWTY with a polynomial fit to the data. Curve d, triangles, dansylated hairpin peptide. Inset, β -hairpin population calculated from curves a, b and c. Curves b and d were obtained from fits with a two-state model (folded and unfolded). The same curves were obtained with the statistical mechanical model.

experimental data, the model produces a funnel-like free energy surface¹⁵ for the β -hairpin peptide (Fig. 4a). The prominent feature of this surface is that it contains only two global minima, and therefore a distinctly bimodal population distribution (two-state behaviour) (Fig. 4c). We identify these populations as the 'unfolded' and 'folded' states of the β -hairpin peptide. In the folded state, 99% of the molecular species contain the complete hydrophobic cluster (Fig. 1), with the most populated species having frayed ends (Fig. 4c). In spite of its kinetic complexity, the model predicts a single exponential relaxation as a consequence of a single overall free-energy barrier (Fig. 4), in agreement with our experimental results (inset to Fig. 3a). The overall transition state should correspond to the peak along the minimal free energy path (labelled $\ddagger$ in Fig. 4b). This is confirmed by the finding that only 1% of barrier crossings use alternative paths of higher free energy. The transition state species represents an ensemble of structures which contains the 45–52 stretch of native peptide bonds and all possible non-native ψ , ϕ pairs for the remaining peptide bonds. This species has a lower energy than the unfolded peptide, explaining the negative activation energy for folding in the two-state analysis. As in proteins, there is roughness in the landscape, with local free-energy barriers up to about half the size of the overall barrier (Fig. 4b).

We have also used the model (without side-chain interactions) to analyse the results obtained with laser temperature-jump experiments on a polyaniline-based helical peptide^{6,7}. The model suggests that the difference between the helix and the hairpin is not in the rate of adding a new native peptide bond to the growing structure ($\sim 1 \text{ ns}^{-1}$ for both), but is in the height of the free-energy barrier separating folded and unfolded states. In the case of helices, the

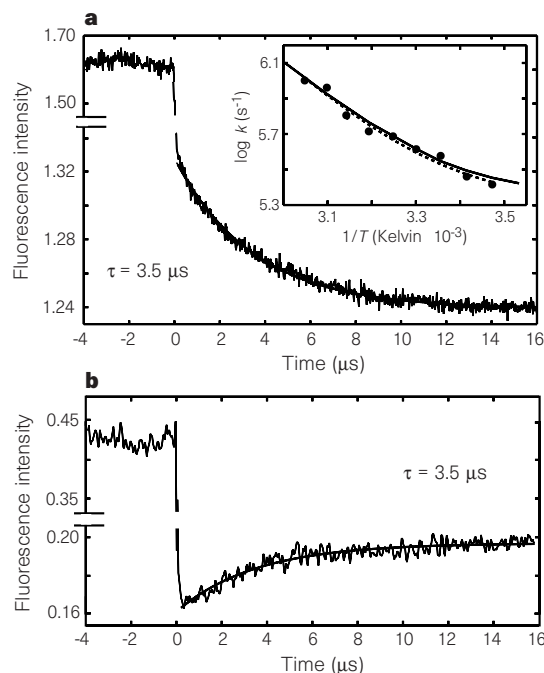


Figure 3 Kinetics of β -hairpin unfolding/refolding. **a**, W43 fluorescence of the hairpin peptide following a temperature jump from 273 K to 288 K. The initial rapid decrease is the result of a change in the intrinsic tryptophan fluorescence due solely to the change in temperature, as shown from experiments on the reference peptide (GEWTY) which shows only this fast phase. The slower phase corresponds to the unfolding and refolding of the hairpin. Inset, Arrhenius plot of the relaxation rate. The circles are the data, the dashed curve is calculated assuming a two-state model, and the continuous curve was produced by the statistical mechanical model. **b**, W43 fluorescence of the dansylated hairpin peptide following a temperature jump from 273 K to 288 K.

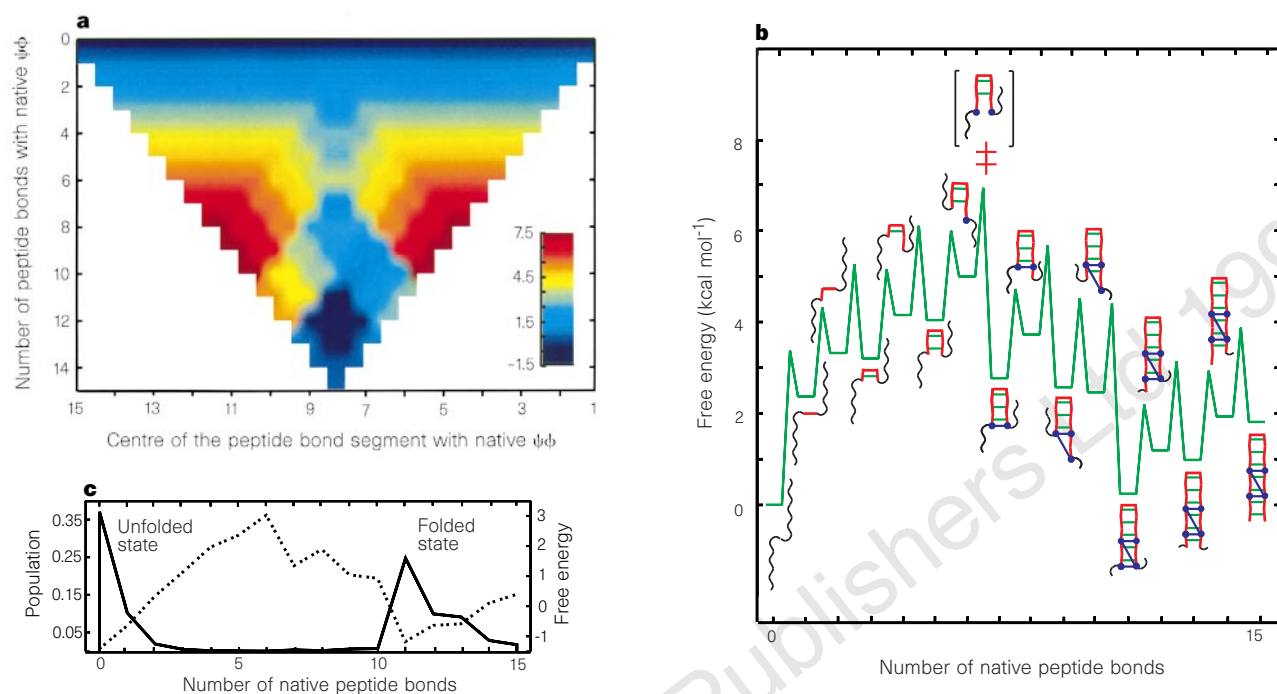


Figure 4 The free-energy landscape of the β -hairpin at the midpoint of the thermal denaturation transition. **a**, Free energy as a function of molecular conformation. Each position on the surface designates the free energy of an equilibrium molecular species, defined by the number of contiguous peptide bonds with native ψ , ϕ angles and the location in the sequence of the central native peptide bond. For clarity, the surface has been smoothed by a linear interpolation. **b**, Free energy as a function of the number of contiguous native peptide bonds along one

of the two minimal free-energy paths for β -hairpin folding. The kinetic barriers between individual species are also shown. Red, green and blue lines signify native peptide bonds, hydrogen bonds and hydrophobic interactions, respectively. **c**, Equilibrium population (continuous curve) and free energy (dashed curve) as a function of the number of contiguous native peptide bonds for the hairpin.

barrier is crossed after fixing three residues in their native conformation to form the first hydrogen bond. Elongation is then thermodynamically favourable for an alanine-based helix because the entropy loss of incorporating another residue (that is, ϕ_i, ψ_i pair) into the helix is more than compensated by the energy from the new hydrogen bond. In contrast, for the hairpin, because two peptide bonds (one ψ_i, ϕ_{i+1} pair in each strand) have to be fixed to form one interstrand hydrogen bond, the process is continuously uphill in free energy until there is stabilization by side-chain interactions. The rate of hairpin formation is therefore sensitive to the position of the stabilizing residues in the sequence (for example, the model predicts that moving the tetrad of hydrophobic residues one residue towards the β turn will speed up hairpin formation by about fourfold). The second important factor in lowering the free-energy barrier and thereby producing faster rates for helices, is the larger number of productive ways of forming helices compared to hairpins. This occurs because a stretch of α -helix may form at any position along the chain with equal probability, whereas a stretch of β -hairpin structure is much less likely if it does not include the central turn region. In simulations of simple bead models using Langevin dynamics, hairpins are also found to fold more slowly, with rates surprisingly similar to the experimental values¹⁶.

Our analysis suggests that the most probable way to initiate hairpin formation is from the β turn. However, this may be too simplistic and other mechanisms could play a role. For example, the semi-empirical estimate of 1 μ s for the formation of a loop of ~ 10 residues¹⁷ suggests that hairpin formation could also be initiated by loop formation to first form the hydrophobic cluster. Finally, our experimental and theoretical analysis exposes the intriguing result that a molecule as small as a 16-residue hairpin exhibits many basic features of protein folding. These include stabilization by both hydrogen bonding and hydrophobic interactions¹⁸, two-state

thermodynamics and kinetics, as found for small proteins¹⁹, and a funnel-like, partially rough free-energy surface, as described by energy landscape theory¹⁵. These findings suggest that the β -hairpin is a suitable system for an in-depth study of fundamental issues in protein folding. □

Methods

Laser temperature-jump apparatus. The laser temperature-jump kinetic spectrometer, described in detail elsewhere^{7,20}, employs a ~ 4 ns Raman-shifted Nd:YAG pulse at 1.54 μ m directly to heat water and an intracavity frequency-doubled argon ion laser operating at 264 nm for excitation of tryptophan fluorescence.

Materials. All peptides were $>95\%$ purity and were purchased from California Peptide Research (Napa, CA). For kinetic experiments the peptide concentration was 400 μ M, which is sufficiently dilute to prevent aggregation at all temperatures studied.

Quantum yields. Quantum yields were determined by comparison with *N*-acetyl-L-tryptophanamide, taken as 0.13 at 298 K. The fluorescence of the reference peptide (GEWTY) (curve c in Fig. 2) becomes identical to that of the hairpin peptide at high temperature, and is assumed to correspond to the fluorescence of the fully unfolded hairpin peptide at all temperatures. The temperature dependence of the quantum yield for the folded hairpin (curve a in Fig. 2) was obtained by multiplying the curve for the reference peptide GEWTY by a constant that best reproduced the relative amplitudes of the fast and slow kinetic phases in the temperature-jump experiments. The fast phase in the hairpin peptide is due to the intrinsic change in tryptophan fluorescence with temperature. Its amplitude is determined by the relative populations of folded and unfolded peptide at the initial temperature, and their quantum yields at both the initial and final temperature. The amplitude of the slow phase is determined by the change in hairpin population upon increasing the temperature and the quantum yields of both the folded and unfolded peptide at the final temperature.

For the dansylated peptide the quantum yields (relative to the undansylated peptide) required to fit the equilibrium data (curve d in Fig. 2) with a two-state model were 0.21 and 0.50 for the folded and unfolded peptide, respectively. This is in reasonable agreement with the values of 0.17 and 0.71 calculated from Forster theory. In this calculation, $r_0 = 2.2$ nm for the tryptophan–dansyl pair, the W43–dansyl distance was estimated as 1.5 nm for the hairpin, and a random-coil distribution with a Flory characteristic ratio of 6 was assumed for the unfolded peptide.

Fitting parameters. The temperature dependence of the observed relaxation rate (inset to Fig. 3) was fitted with a two-state model using the equation: $\ln k_{\text{obs}} = \ln[k_f + k_u] = \ln[A \exp(-E_a/RT)(1 + 1/K)]$ with $A = 2.9 \times 10^4 \text{ s}^{-1}$, $E_a = -1 \text{ kcal mol}^{-1}$, and K obtained from the equilibrium data. The parameters of the statistical mechanical model used in constructing the free energy surface in Fig. 4a, as well as the curves in Figs 2 and 3, were obtained by simultaneously fitting the kinetic and equilibrium data. The resulting parameters (per bond or interaction) are $\Delta S_{\text{conf}}(\text{hairpin}) = -3.2 \text{ cal mol}^{-1} \text{ K}^{-1}$, $\Delta H_{\text{hb}} = -1.1 \text{ kcal mol}^{-1}$, $\Delta G_{\text{sc}} = -2.1 \text{ kcal mol}^{-1}$, $k_0 = 1.5 \times 10^{10} \text{ s}^{-1}$, and $E_0 = 1.0 \text{ kcal mol}^{-1}$. The value of $3\Delta G_{\text{sc}}$ corresponds to the burial of about $300 \text{ \AA}^2$ of nonpolar surface area in forming the 3 interactions in the 4-residue hydrophobic cluster^{21,22}. It is smaller than the value of $420 \text{ \AA}^2$ calculated from the three-dimensional structure, possibly because this calculation neglects the decrease in entropy associated with immobilizing the 4 side chains in the cluster. The same parameters were used for fitting the data on the helix^{6,7}, except that $\Delta S_{\text{conf}}(\text{helix}) = -2.90 \text{ cal mol}^{-1} \text{ K}^{-1}$. With these parameters, the fundamental rates for the single steps of making and breaking a hydrogen bond are, respectively: $k_0 \exp[(T\Delta S_{\text{conf}} - E_0)/RT] \approx k_0 \exp[(-E_0 + \Delta H_{\text{hb}})/RT] \approx 1 \text{ ns}^{-1}$. This value is consistent with the finding by time-resolved infrared spectroscopy that a decrease in β -structure hydrogen bonding in ribonuclease A occurs at 1–5 ns after a temperature jump²³.

Received 21 May; accepted 18 August 1997.

1. Eaton, W. A., Muñoz, V., Thompson, P. A., Chan, C.-K. & Hofrichter, J. Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* **7**, 10–14 (1997).
2. Zimm, B., Doty, P. & Iso, K. Determination of the parameters for helix formation in poly- γ -benzyl-L-glutamate. *Proc. Natl Acad. Sci. USA* **45**, 1601–1607 (1959).
3. Gruenewald, B., Nicola, C. U., Lustig, A. & Schwarz, G. Kinetics of the helix-coil transition of a polypeptide with non-ionic side groups, derived from ultrasonic relaxation measurements. *Biophys. Chem.* **9**, 137–147 (1979).
4. Chakrabarty, A. & Baldwin, R. L. α -Helix stability. *Adv. Prot. Chem.* **46**, 141–176 (1995).
5. Muñoz, V. & Serrano, L. Helix design, prediction and stability. *Curr. Opin. Biotech.* **6**, 382–386 (1995).
6. Williams, S. *et al.* Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **35**, 691–697 (1996).
7. Thompson, P. A., Eaton, W. A. & Hofrichter, J. Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a “kinetic zipper” model. *Biochemistry* **36**, 9200–9210 (1997).
8. Blanco, F. J. *et al.* NMR evidence of a short linear peptide that folds into a β -hairpin in aqueous solution. *J. Am. Chem. Soc.* **115**, 5887–5888 (1993).
9. Kobayashi, N., Yoshii, H., Murakami, T. & Munekata, E. *Peptide Chem.* 278–290 (1993).
10. Blanco, F. J., Rivas, G. & Serrano, L. A short linear peptide that folds into a β -hairpin in aqueous solution. *Nature Struct. Biol.* **1**, 584–590 (1994).
11. Searle, M. S., Williams, D. H. & Packman, L. C. A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-stable non-native β -hairpin. *Nature Struct. Biol.* **2**, 999–1006 (1995).
12. de Alba, E., Jimenez, M. A., Rico, M. & Nieto, J. L. Conformational investigation of designed short linear peptides able to fold into β -hairpin structures in aqueous solution. *Folding & Design* **1**, 133–144 (1996).
13. Ramirez-Alvarado, M., Blanco, F. J. & Serrano, L. De novo design and structural analysis of a model β -hairpin peptide system. *Nature Struct. Biol.* **3**, 604–611 (1996).
14. Sieber, V. & Moe, G. R. Interactions contributing to the formation of a β -hairpin-like structure in a small peptide. *Biochemistry* **35**, 181–185 (1996).
15. Bryngelson, J. D., Onuchic, J. N., Socci, J. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape theory of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).

16. Klimov, D. K. & Thirumalai, D. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* **79**, 317–320 (1997).
17. Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc. Natl Acad. Sci. USA* **93**, 11615–11617 (1996).
18. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Prot. Chem.* **14**, 1–63 (1959).
19. Fersht, A. R. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3–9 (1997).
20. Thompson, P. A. in *Techniques in Protein Chemistry* Vol. VIII (ed. Marshak, D. R.) 735–743 (Academic, San Diego, 1997).
21. Baldwin, R. L. Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl Acad. Sci. USA* **83**, 8069–8072 (1986).
22. Livingstone, J. R., Spolar, R. S. & Record, M. T. Jr Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry* **30**, 4237–4244 (1991).
23. Phillips, C. M., Mizutani, Y. & Hochstrasser, R. M. Ultrafast thermally induced unfolding of RNase A. *Proc. Natl Acad. Sci. USA* **92**, 7292–7296 (1996).
24. Gronenborn, A. M. *et al.* A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657–661 (1991).

Acknowledgements. We thank A. Szabo and P. Wolynes for helpful discussions. We also thank J. Omichinsky for providing material used in the preliminary stages of this work. V.M. was supported by a postdoctoral fellowship from the Human Frontiers Science Program Organization.

Correspondence and requests for materials should be addressed to either V.M. or W.A.E. (e-mail addresses: vmunoz@helix.nih.gov; eaton@helix.nih.gov).

correction

K⁺ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons

Dax A. Hoffman, Jeffrey C. Magee, Costa M. Colbert & Daniel Johnston

Nature **387**, 860–875 (1997)

On page 875 of this Article, there are some typographical errors in the equations used in the computer model. The errors are in the third and fourth full paragraphs on this page.

The first sentence of paragraph 3 should begin: “The Na⁺ channel model was of the form, $g_{\text{Na}} = 4.2\bar{g}_{\text{Na}}m^3h$, ... The equation for α_m in paragraph 3 should read: $\alpha_m(\nu) = 0.182(\nu + 32.5)/(1 - \exp(-(\nu + 32.5)/4.5))$.”

The first sentence of paragraph 4 should read: “The A-type channel models were of the form $g_{\text{KA}} = \bar{g}_{\text{KA}}m^4h$.” The equation for β_h in paragraph 4 is mislabelled as β_m and the α_h equation has an incorrect minus sign. The sentence should read: “ $h_{ss}(\nu) = \alpha_h(\nu)/(\alpha_h(\nu) + \beta_h(\nu))$, where $\alpha_h(\nu) = -0.01(\nu + 58)/(\exp((\nu + 58)/8.2) - 1)$ and $\beta_h(\nu) = 0.01(\nu + 58)/(\exp((\nu + 58)/-8.2) - 1)$ for all A-channels.” □

Passion and prejudice in research

In every country examined, there are far fewer women in senior positions in science than the number who begin a career in research – and those who do achieve success are paid less than their male colleagues. Irrespective of gender, the distribution of grants for scientific research is grossly skewed in favour of the few at the expense of the many. And members of some ethnic minority groups tend not even to begin careers in research. Why are these discriminatory practices tolerated, and what (if anything) is being done to change this state of affairs?

Women researchers are rare elements in Germany

Scot Stevenson

A strange new form of plural is circulating in German universities. The plural of male professors, *Professoren*, which is also the generic plural for groups of men and women, has been merged with the plural for female professors, *Professorinnen*, to form *ProfessorInnen*, spelled with a capital I. This construction, which has become fashionable in some circles, allegedly creates a friendlier climate for women.

Some may wince at this example of political correctness, but Germany is one of the industrial nations with the lowest percentage of women holding senior positions in teaching and research. A UNESCO study in 1996, for example, placed Germany behind Portugal, Korea and Mexico on the number of women in physics faculties.

The government is trying to change this situation. In 1985 it amended the *Hochschulrahmengesetz*, the law giving general guidelines to the *Länder* (regional governments) responsible for the universities, to require them to specify equal opportunities for women scientists. And in 1989, the Bund-Länder Kommission, which co-ordinates federal and *Länder* research policy, provided recommendations for universities and research centres to promote women in science and research.

These measures have forced some progress. The number of first-year female university students has exceeded the number



Cherchez la femme: annual meeting of biomedical section heads of the Max Planck Society, Bremen, 1997.

of male students for the past two years. Every German university has a *Frauenbeauftragte*, an official in charge of women's affairs, although their authority and funding varies from state to state. And the Deutsche Forschungsgemeinschaft, Germany's largest research grant-giving body, increased the number of grants to women to achieve *Habilitation*, a qualification required to become a professor, from 21 in 1990 to 136 in 1996.

At higher ranks, however, the percentage of women drops sharply. At the most senior level, that of a 'C4' professor, only 4.8 per cent of university positions are held by women.

Things are no better at research institutes outside universities. In 1995, only 2.3 per cent of top research positions were held by women, and there were no female directors at all. The percentage of women at the equivalent level to C4 professors at Max Planck research institutes has remained at about 2.4 per cent for the past 10 years.

Though this problem is not confined to

Germany, it is more extreme here than in other countries. Women in Turkey have a better chance of being promoted to senior academic positions than women in Germany, even though only two-thirds of Turkish women are literate, according to a 1996 study from the University of Kassel.

Various cultural factors combine to make Germany less ready to include women in research. "In old [West] Germany, technology and science were connected to masculinity," says Dagmar Heymann, a speaker for *Frauen in Naturwissenschaft und Technik*, a feminist group that campaigns for more influence for women in science. Things were different in the former East Germany, where in 1995 there was still a higher percentage of women at all levels in science.

As many German university professors are nearing retirement — about half of all positions will be up for replacement by 2005 — many see this as an opportunity to bring more women into research.

In July, the Bund-Länder Kommission released a five-point-programme with the aim of considerably increasing the number of women scientists. The programme suggests, rather vaguely, "competition between institutes" as one means to reach this goal.

German science minister Jürgen Rüttgers says he wants 20 per cent of high-ranking positions to be occupied by women by 2005. To this end, 20 per cent of the DM3.6 billion (US\$2 billion) budget of the new *Hochschulsonderprogramm III*, a package of special funds for universities, is to be reserved to promote these goals. Politicians of the opposition SPD and Green parties, which could come to power after the next general

Remaining at the bottom of the UK league

Although women account for half the students taking degrees at most UK universities, they held only 8 per cent of professorships in 1995–96, according to figures collated by the UK Higher Education Statistics Agency.

The picture is worst for women in engineering and technology: of 981

professorships in the United Kingdom in these disciplines, only 14 were held by women (a meagre 1.4 per cent); there were no women professors in agriculture. Subjects connected with medicine were a little better, with women holding 19.9 per cent of chairs.

The most prestigious UK

universities are among the poorest performers. Oxford has 15 women professors out of 232; Cambridge has 13 out of 247. But the newer universities have made a concerted effort to crack the glass ceiling — 30 per cent of the South Bank University's chairs, for example, are held by women.

H.Q.

elections in 1998, say they would go further and link funding to progress in equal-opportunity programmes.

Individual *Länder* have started their own support programmes. Nordrhein-Westfalen initiated the Lise Meitner programme in 1991 to fund women in their *Habilitation* for up to three years. But, despite spending DM15.6 million (US\$9 million), the number of female C4 professors there is still at 5.8 per cent. "Progress is a snail," says Nordrhein-Westfalen's science minister, Anke Brunn.

Frustrated with the slow progress at federal and state levels, the Free University of Berlin set up an incentive system in 1995 whereby each faculty pays 1 per cent of its budget into a common pool. The money is redistributed to the faculties according to a rating system that takes into account the promotion of women compared with the previous year. If a faculty's score has improved, it receives more money than it paid into the pool, but if it did worse, it receives less.

Brigitte Mühlenbruch, speaker for the Federal Conference of Germany's *Frauenbeauftragte*, says that an amended version of the Berlin system could be a model for all German universities. But she says it is important that it should be based on rewards only, not punishments, and the amount of money should be raised considerably, to "5 or 10 per cent" of the faculties' budgets. She would like to see progress on equal opportunities become a basic criterion for distribution of funds at all levels, as important as scientific excellence.

The Wissenschaftsrat, Germany's influential science council, has begun an investigation on equal opportunities. Its recommendations, expected in the new year, will have a major impact. Although nobody expects the council to propose radical changes, such as quotas or positive discrimination, one thing is sure: *ProfessorInnen*, with or without its capital I, may expect more than a mere change of language. □

Scot Stevenson is a student of journalism at the Johannes Gutenberg University in Mainz.

Along the leaky pipeline

Potter Wickware

US women drop out of research more often than men, so that by the time they reach the end of the 'leaky pipeline' — from undergraduates through to faculty appointments — they are substantially under-represented. By contrast, ethnic minorities — African Americans, Hispanics and American Indians — tend not even to begin science studies.

This is not only a squandering of intellectual and creative resources, but is a contradiction of the popular notion that science ought to be disinterested, open to all, and free of the taint of discrimination.

According to the National Science Foundation, women account for only 22 per cent of scientists and engineers in the US labour force. African Americans, Hispanics and American Indians account for only 6 per cent of the science and engineering labour force, although collectively these groups amount to more than a quarter of the population.

These proportions decline even further on graduation — there are fewer women and minorities in the science and engineering labour force than obtained science and engineering degrees. But there are rays of hope: since 1990, 32 per cent of scientists and engineers who received degrees are women and 8 per cent are black, Hispanic or American Indian. Nevertheless, there is concern that it will be hard to attain a representative demographic distribution given the difficult times for scientific research at present.

John Matsui, who directs the biology scholars programme at the University of California, Berkeley, a support group for minority students, believes that two obstacles stand in the way of success for under-represented individuals in science. One is the traditional nature of introductory chemistry, maths, biology and physics courses.

"These courses are not designed with the idea of retaining students, particularly minority students, who may be the first in their families or immediate communities to go to college," he says. These students perhaps also lack personal models of academic achievement, and may as well be engaged in economic, cultural or other struggles not familiar to the typical undergraduate of the past. Introductory courses continue to follow the educational model that permitted faculty to excel in their own undergraduate careers. It is a model that continues to select for and reproduce successful students who learn and respond like the faculty. The effect, although probably not intentional, is ultimately one of discrimination against people from disparate backgrounds, says Matsui.

The second hazard to minority achievement in science, he continues, is a lack of explicit connection between introductory course content and its relevance to career paths. "Typically, these courses don't address questions of the students' 'goal ambiguity', or the pros and cons of scientific careers. 'What does a biologist do?' 'What use is a biology major?' are largely unanswered questions." Although minority students are not alone in wrestling with these doubts, majority students are perhaps better equipped to resolve them, because they are more likely to have examples in their lives of people — teachers, family or friends — who have successfully completed their studies.

In the past, undergraduates were perhaps better able to tolerate the apparent lack of relevance of demanding introductory science courses, and to take the answers to such questions on faith, because they were prepared for the rigours of scholarship by values absorbed from social class and family. But under-represented students, new to college, may well lack the support imparted by these value systems. They are at risk of failing to persevere in the introductory courses.

Women, as they find themselves bumping up against the glass ceiling, are ostensibly in a different situation from those in ethnic minorities. But the problem is traceable to same root cause: a parallax shift of viewpoint from instructor to student. "Men and women have their own ways of doing science," observes Caroline Kane, a biochemist at the University of California, Berkeley. "There's a tendency for males to be competitive hotshots, with a 'let's beat them' mentality. Women tend to be more collaborative."

The traditional organization of many faculties militates against cooperation in favour of a more cutthroat, individualistic style, simply because university science faculties tend to be dominated by white males.

In the context of a PhD glut, static levels of undergraduate enrolments and federal research support, in a time when science is entering what Arthur Kornberg has called its "stationary phase", why should science be

Waiting for a head count

One problem facing science policy makers both in Germany and across Europe is a lack of statistical data on under-represented groups. In Germany, some information simply is not collected. The Deutsche Forschungsgemeinschaft, for example, does not record how many of its grants go to researchers with specialist needs. The Bund-Länder Kommission began to collect data on appointment procedures only

this year, registering the number of women who apply for a professorship, as well as the number that are finally accepted.

Germany does not have a Freedom of Information Act, and has strengthened privacy laws over the past two decades to the point where certain statistics have become impossible to collect.

The European Commission's Training and Mobility of Researchers

programme, which provides fellowships for young scientists to do research in other European Union countries, is accepting bids for a study on women in science. Data collected from the 15 member states and four affiliated countries will also be used to examine the participation of women recipients of the programme's grants. The data should become available at the beginning of 1999.

S.S.

concerned about opening up to a larger instead of a more restrictive pool of applicants? The all-important reason, says Kane, is diversity of ideas. "Everybody brings their own experience to science. People's life experiences and cultural antecedents are different, sometimes radically so, and these will affect choice of research problems. The result will be an expansion of creative breadth."

How to bring about the necessary changes? Shirley Malcom, director of education and human resources for the American Association for the Advancement of Science, and an adviser to President Bill Clinton on science issues, stresses that increased support must be provided from pre-college through to faculty stages. Structural change in higher education must include stronger emphasis on changing introductory science courses, increased interaction between students and professionals — as well as professors — career advice and mentoring, and undergraduate research opportunities. □

Potter Wickware is a science writer in Oakland, California, USA. e-mail: wick@netcom.com

In defence of small science

Frederick Sachs

It's 1904. Einstein gets three grants. *Annus mirabilis* becomes *annus satisfactory*. It seems obvious, from our present perspective, that for Einstein to write and administer a collection of grants would have destroyed his creativity. The US National Institutes of Health (NIH), however, supports many researchers who are principal investigators of four, five, six or more research grants. Have these grants been awarded without concern for the applicant's scientific creativity? What peer-review committees (called study sections) make these awards? Although giving many grants to one person probably doesn't often result in bad science (although a lack of supervision leading to fraud may be an issue), the real penalty is the loss of ideas. We are giving funds to provide predictable new life at the expense of the unexpected¹.

Given a fixed NIH budget, one person can only get their research funded at the expense of someone else. For every grant awarded to someone who already has other NIH grants, one or more people are dismissed. Instead of multiple grants requiring less money because they build on the infrastructure of pre-existing grants, the average value increases with the number of grants held by one person^{2,3}. Needless to say, the losers are most likely to be beginners who do not have an existing laboratory to generate preliminary data. A competitive NIH grant submission has about a 90 per cent chance of getting rejected^{4,5}, and this high probability encour-

ages smart young people to leave research.

In 1986, more than 96 per cent of grant awards in the category reserved for first-time applicants went to people under 36 years of age. By 1993, this proportion had fallen to less than 29 per cent (ref. 5). In 1994, 38 per cent of all first-time awards went to people between 36 and 40 years old⁶. This trend is bad for science. Biology has made its greatest advances from the ideas of young independent investigators. The shortage of funds has its greatest effect on those who are not 'in', as shown by that ground-breaking study⁷, done in Sweden, showing that women needed vastly superior credentials, equivalent to two or three extra publications in *Nature* or *Science*, to have the same chance as men of getting a postdoctoral fellowship (see page 204). The study showed that men who were not associated with members of the peer-review committees were also treated unfairly.

There is no reason to believe that the United States is different. In 1995, 45 per cent of PhDs in biomedical sciences (US citizens in US universities) were awarded to women, but less than 8 per cent of all research-centre grants were awarded to women^{6,8}. How many women seek funding? Not as many as men, if the Federation of American Societies for Experimental Biology can be taken as representative of the funded biomedical research community⁹: women constitute only about 13 per cent of its membership.

The NIH appears to be making an effort to maintain balanced reviews: its study sections contain on average 20 per cent women. Coincidentally, about 20 per cent of research projects are awarded to women. The presence of a male majority on study sections, however, could be a significant factor in rating proposals from women. This is particularly important now because the number of female graduates is up from 25 per cent in 1975 to 45 per cent in 1995.

How many grants should a person control? A random search of the CRISP database (<http://www.nih.gov/grants/award/crisp.htm>) shows at least one principal investigator on 11 different grants, some of which are large programme project grants. Because CRISP is not relational, it is not possible to identify the real grant winners. CRISP also has no information on whether a principal investigator is a co-principal investigator on any other NIH grants, a recipient of grants from other sources, or has any administrative or clinical duties — or even for how much the grants are funded.

The principal investigator with 11 grants, however, has less than half a day to work on each one during a five-day week. This assumes no travel, and no institutional obligations. Should this level of commitment be encouraged while other applicants are consigned to the trash? If we give someone US\$2 million a year to perform a clinical study, should we give them six more research

grants (NIH lists some people as principal investigators on seven such grants²)? Should we add to that grant supply some training grants that pay for students to work on these projects? In large laboratories most training is done by other students and fellows, not laboratory directors. In compiling data on multiple grants, I received no responses to questions about justification for funding from some of these successful principal investigators. Perhaps they were too busy.

The publicly funded research enterprise is short of money to support independent investigators. Where should it come from? As things stand, it can come either from an increased NIH budget or from a redistribution. (A large increase in funding from Congress seems unlikely, and in any event, given present funding policy, an increased budget would lead to more money going to currently funded investigators.) If, instead, the NIH were to limit the number of grants and to decrease the overhead rates for multiple grants from the same principal investigator, there could be significant changes without a decrease in the number of funded labs. Limiting the number of grants to two for each principal investigator would provide funds for 3,000 new principal investigators. One thing is clear: the NIH must act now to stop the loss of promising young investigators. □

Frederick Sachs is in the Department of Biophysical Sciences, State University of New York, Buffalo, New York 14124, USA.

e-mail: sachs@fred.med.buffalo.edu

1. Jones, S. *The language of the genes: biology, history and the evolutionary future* (Doubleday, New York, 1993).
2. Sachs, F. *Nature* **388**, 222 (1997).
3. Sachs, F. *Nature* **363**, 578 (1993).
4. Mandel, H. G. *Science* **269**, 13–14 (1995).
5. Ruzek, D. L., O'Neil, E., Williard, R. & Rimel, R. W. *Trends in US funding for Biomedical Research*. (UCSF Center for the Health Professions, San Francisco, 1996).
6. *NIH extramural trends* (NIH Publication No. 96–3506, 1995).
7. Wennerås, C. & Wold, A. *Nature* **387**, 341–343 (1997).
8. Chandra, S. et al. *Proc. Natl Acad. Sci. USA* **86**, 1870–1874 (1989).
9. Garrison, H. & Heinig, S. J. *FASEB J.* **9**, 703–706 (1995).

"Invasion of dragons" needed

Helen Gavaghan

"Nature herself prescribed to the woman her function as mother and housewife and that law of nature cannot be ignored... without grave damage which... would especially manifest itself in the next generation." So wrote Max Planck in 1897. It would be hard to find a more discriminatory statement.

Planck was hampered by the social and scientific mores of his time, but what of the prospects for women in physics in our more enlightened days? They consistently fare less well in the physical than the biological sciences where, argued Gerhard Sonnert and Gerald Holton of Harvard University in *The*

American Scientist (January–February 1996), “the gender gap has all but disappeared”. In a study of men and women in physics, maths and engineering, Sonnert and Holton found there was a significant gap between academic ranks attained in the United States. The gap was particularly pronounced in younger scientists.

Although the situation is slowly becoming more equitable (women account for 15 per cent of PhDs in the United States compared with 3 per cent 30 years ago), the disparity between the numbers of women and men succeeding in physics persists.

In the July 1997 issue of *Physics and Society*, for example, Jolanta Lagowski and Janis McKenna found that 18 per cent of those receiving a bachelor's degree in physics in Canada were women; 13 per cent of all PhDs in physics went to women and 5 per cent of faculty staff were women. Only 2 per cent of tenured physics faculty members were women. The authors collected responses from 40 institutions, and of these 80 per cent had one or no women faculty members in their physics departments. Nearly half of the 40 institutions had no women faculty members in physics. The figures, say Lagowski and McKenna, are similar to those in the United States but worse than in Europe. In France, Italy and Turkey, 23 per cent of physics faculty members in 1991 were women.

Although Lagowski and McKenna do not attempt to explain the differences they found, D. Elizabeth Pugel in an article in the same issue explores the stages from birth in the development of a physicist. A sociological approach highlights four factors in the nurturing of a scientific outlook: parental behaviour, toy selection, and the wider relationships of the child and the adolescent.

For nearly 30 years, researchers have noted that most people allow infant boys greater freedom than girls to crawl around exploring their surroundings. Stereotyping follows, which can, of course, be reinforced by the toys that parents select. Pugel points out that the selection of dolls that purportedly enhances social skills is not a bad thing in today's large collaborative teams of physicists. The problem occurs when these social skills are perceived as drawbacks by others.

Some school teachers, says Pugel, are still of a generation in which a science education for girls was not particularly emphasized. Even if they do teach science to young children, women teachers may be intimidated by science and provide poor role models. Few, if any, women physicists have the public status of Einstein or Oppenheimer.

Nevertheless, young girls do become physicists, even though the data show them trapped firmly beneath the glass ceiling. In the July 1996 issue of *Physics and Society*, Howard Georgi, former chairman of the physics department at Harvard University, wrote of senior faculty meetings at Harvard:

Equality not taken for granted

The response across the world could be measured on the Richter scale after the revelation that the Swedish medical research council (MRC) exercised prejudice in its allocation of research fellowships (A. Wold and C. Wennerås, *Nature* **387**, 341–343; 1997).

Six months later, the implications are still being discussed in the newspapers and on radio and TV. But has anything really changed? Most definitely yes, say Agnes Wold and Christine Wennerås, the authors of the *Nature* Commentary. The MRC has finally accepted that it had been acting unfairly, and has changed its procedures. And some research councils in other countries – for example, the United Kingdom (see Breen, G. *Nature* **389**, 326; 1997) – are checking their own procedures.

Wold and Wennerås had to fight hard to convince Sweden's MRC that it had a problem. They began their investigations into its peer-review system two years ago, but were hampered by its lack of cooperation. The research council's belief in its system of meritocracy was unshakeable, says Wold. She and Wennerås had to get

court orders to force the MRC to make documents available. Their analysis of the peer-reviewers' reports showed that in 1995 (the only year they were able to study), a woman applying for a postdoctoral fellowship had to be two-and-a-half times more productive than a man to rate the same scientific competence scores by referees. The analysis revealed that connections to any of the reviewers, independent of gender, helped bump up competence scores.

Wold believes that the MRC was genuinely unaware of the prejudices it was harbouring. Women reviewers were not significantly fairer than men, she says, when it came to estimating the skills of their own sex.

After the *Nature* article was published, the MRC began its own studies into possible prejudice in its allocation of project money, which it says it intends to publish. Although it found no evidence that the scientific competence of women had been misjudged, it did find that it allocated smaller grants to women than to men with identical competence ratings.

Jan Nilsson, vice-secretary

of the MRC, says that he and the MRC's subcommittees which had advised on grant distribution were shocked by the revelation. “It came as a complete surprise. We had a fair system for grading scientists but transforming the grading to size of grant proved – unexpectedly – to be less rigid”. The MRC has already corrected this tendency, he says, and the size of grants allocated this autumn were based only on competency scores. This proves, says a delighted Wold, that there is no truth in the adage that “things can only change slowly because they have been like this for hundreds of years”.

Wold is intolerant of attempts to personalize the issue of discrimination. Discrimination is simply about prejudice, she says, and has nothing to do with family status, self-esteem, or any other fantasized female attribute that some claim contributes to women being taken less seriously. “It is a purely statistical problem and making it personal serves only to lower standards of discussion.” Wold has now raised the discussion to a level that has brought tangible results.

Alison Abbott

“I was appalled by the old-boys-club atmosphere that oozed from these gatherings, and I began to feel that an invasion of dragons was needed to shake up the country club.”

Georgi believes that women often seem outwardly less sure of themselves than men. Initially, he found that this diffidence made it difficult for him to communicate as effectively with women as with male students. He also consistently discovered that women graduate students were more talented than was suggested by their entry examinations.

This point, argues Georgi, suggests the need for an affirmative-action programme — an unfashionable view both in academic institutions and in the law courts (see *Nature* **376**, 288; 1995 & **384**, 97; 1996). Thus if there is reason to suspect that woman or ‘minority’ students are better than their applications suggest and an interview reveals them to be as capable as the best white male students, then there is a case for affirmative action. Of those women that do gain faculty positions,

Georgi argues, “We are still shoe-horning women into a programme that works well for men but not for women and then trying to deal with the problems that arise.” Georgi concludes: “You have to be an optimist and keep trying... Things are getting better, but always more slowly than we would like.” □

Helen Gavaghan is a freelance science and technology writer based in Hebden Bridge, UK.

Careers and recruitment in *Nature*

Earlier Careers and Recruitment features include:

- **Immunology** (13 February issue)
- **Plant science** (22 May issue)
- **Bioinformatics** (25 September issue)

Nature welcomes comments from readers, in the form of information about recruitment programmes, readers' experiences, or their reactions to the articles published. Comments should be sent in the first instance to:

Maxine Clarke at m.clarke@nature.com.